REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2005 Name of Author GALGI C-V

Genotypic and Phenotypic Variation in the Human Immunodeficiency Viruses

Catherine Victoria Gale

Submitted to the University of London for the degree of Doctor of Philosophy

May 2005

Centre of Virology
Department of Infection
Windeyer Institute of Medical Sciences
University College London

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Acknowledgements

There are many people that I should thank for making the completion of this thesis possible. For reagents thanks to Greg Towers, Stuart Neil, Deenan Pillay, Jim Hoxie and Hiro Hoshino, for the cDNA arrays Tom Freeman and Debbie Williams (HGMP, Cambridge) and for Cat3 lab facilities, Robin Weiss. Thanks to Liz Thompson for keeping me organised and to David Marchant for keeping me in one piece, when I started out in the Wohl.

For his supervision and support, through some pretty horrible times, thanks to Paul Kellam. And of course everyone else who mopped/propped/cheered me up on said occasions – you know who you are and how much your support has meant over the last 3 years. Special mention goes to Rich Myers, without whom Chapter 1 of this thesis would not exist, and with whom I've really enjoyed working.

To my mum and brother – thanks for your support and encouragement. And to my Dad, whilst I blame you entirely for getting me into this whole science thing, I wish you could be here and I hope you would be proud. And finally thanks to Sam Wilson. For everything.

This PhD was funded by efforts of Richard Tedder and consequently the Special Trustees at the Middlesex Hospital.

Abstract

Despite the involvement of multiple genetic variants of HIV in the causation of disease worldwide, most research has focussed on subtype B, the prevalent subtype in the western world. As a consequence, it is currently not clear whether genetically distinct HIV strains have different biological properties that cause differences in *in vivo* transmission, disease progression, replication capacity or sensitivity to antiretroviral drugs. In order to increase the ease of classification of viral diversity and thus aid studies into its importance, a novel genotyping tool for classifying HIV-1 subtype based on *pol* sequence, produced routinely during drug resistance monitoring, has been developed in this thesis. A dataset of 187 full-length HIV-1 sequences was used to generate Gag, Pol, Protease-Reverse Transcriptase (PR-RT) and Env protein sequence alignments. Phylogenetic analyses enabled generation of subtype specific alignments and, whilst sequence variation in the PR-RT dataset was low, this variation was adequate for PR-RT subtype assignment. The subtyping tool, named STAR, utilises position specific scoring matrices (PSSMs) derived from these subtype specific multiple sequence alignments and results in highly accurate reclassification of the subtype alignment sequences, with 98.6% of sequences being accurately assigned a subtype.

Subsequent to the development of STAR the importance of HIV genetic variation classified as subtype, was addressed. A comparison of the relative growth capacity of HIV-1 primary isolates of subtypes A, B, C, D, F, group O and HIV-2 was performed in two T-cell environments. A novel reporter cell line was developed specifically to facilitate this work. Clear and consistent differences in *in vitro* growth phenotype in terms of rate and cytopathogenicity were detected, indicative of intrinsic differences between the HIV-1 types and subtypes. This work was extended by the utilisation of microarray technology which offers the possibility to analyse, at any given time point, the transcriptome of a virus-infected cell. A comparison of the transcriptional responses within T-cells to infection with HIV-1 subtype B, Group O and HIV-2 enabled identification of both core and diverging transcriptional response programs. Whilst the core response program provides insight into the most essential interactions between virus and host during HIV infection of T-cells, analysis of the diverging responses provide evidence that genetically divergent strains of HIV may interact differently with the host. It is proposed that these differences may have the potential to influence disease outcome.

Contents

Figures

Chapter 1.0

Chapter 2.0

Chapter 3.0

Chapter 4.0

Tables

Chapter 1.0

Chapter 2.0

Chapter 3.0

Chapter 4.0

Abbreviations

AIDS	Acquired Immune Deficiency Syndrome
AIP1/ALIX	ALG-2-interacting protein X
AP-1	Activator protein 1
AP-2	Adaptor-related protein complex 2
AP-2 gamma	Activating enhancer binding protein 2 gamma
AP	Alkaline phosphatase
APC	Antigen presenting cell
APOBEC	Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like 3G
APV	Amprenavir
ARNA	Amplified RNA
ART	Antiretroviral therapy
ARV	AIDS-related viruses
ATP1G1/HOMG2	FXYD domain containing ion transport regulator 1
AZT	Azidothymidine
BAD	BCL2-antagonist of cell death
BCL-6	B-cell lymphoma-6
BIRC3	Baculoviral IAP repeat-containing 3
BLAST	Basic local alignment search tool
Bp	Base pairs
CCD	Charge coupled device camera
CCR5	CC chemokine receptor-5
Cdk5	Cyclin dependent kinase-5
CLC-4	Chloride channel-4
COP	Non-clathrin-coated vesicular coat
CPE	Cytopathic effects
Cpx	Complex
CRF	Circulating recombinant forms
CTD	C-terminal domain
CTL	Cytotoxic T-cell
cDNA	Copy DNA
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
CDC	Centres for Disease Control
CDK	Cyclin dependent kinase
CS	Cleavage sites
CTCF	CCCTC-binding factor
CVG	CEM-GFP-R5
CypA	Cyclophilin A
CXCR4	CXC chemokine receptor-4
DC	Dendritic cell
DC-SIGN	DC-specific intercellular adhesion molecule (ICAM)-grabbing non-integrin
d	d-value or relative difference
dATP	Deoxyadenosine triphosphate
DAVID	Database for Annotation, Visualization and Integrated Discovery
dCTP	Deoxycytidine triphosphate
ddATP	Dideoxyadenosine triphosphate
ddCTP	Dideoxycytidine triphosphate
ddGTP	Dideoxyguanosine triphosphate
ddUTP	Dideoxyuridine triphosphate
ddI	Dideoxyinosine
DEPC	Diethylene pyrocarbonate
DHCR	Dehydrocholesterol reductase
dITP	Deoxyinosine triphosphate
dTTP	Deoxythymidine triphosphate
DMEM	Dulbecco's modified eagles medium
DNA	Deoxyribonucleic acid

DNAJA2	DnaJ (Hsp40) homolog, subfamily A, member 2
dNTP	Deoxynucleotide
dsRNA	Double stranded RNA
EBV	Epstein barr virus
EDTA	Ethylenediaminetetraacetic acid
EFV	Efavirenz
EGFR	Epidermal growth factor
EGR	Early growth response
env	Envelope
ER	Endoplasmic reticulum
ERV	Endogenous retrovirus
FACS	Fluorescence activated cell sorter
FASTK	Fas-activated serine/threonine kinase
FBS	Foetal bovine serum
FDR	False discovery rate
FFU	Focus forming units
FGF2	Fibroblast growth factor 2
FN	False negative
FOS	v-fos FBJ murine osteosarcoma viral oncogene homolog
FP	False negative
g	Gravity
GADD45	Growth arrest and DNA damage-inducible protein 45
Gag	Group specific antigen/group associated genes
GALF	Gut-associated lymphoid tissue
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
GO	Gene ontology
gp	Glycoprotein
Grb10	Growth factor receptor-bound protein 10
GTPase	Guanosine triphosphatases
HAART	Highly active antiretroviral therapy
HBV	Hepatitis B virus
HCV	Hepatitis C virus
HCC	Hepatocellular carcinoma
hGNC5	General control of amino-acid synthesis 5-like 2
HIV	Human immunodeficiency virus
HIV-1	Human immunodeficiency virus type 1
HIV-2	Human immunodeficiency virus type 2
HIVAN	Hiv-associated nephropathy
HIVE	HIV encephalopathy
HLA	Human leukocyte antigen
HMA	Heteroduplex mobility assay
HMG CoA	3-hydroxy-3-methylglutaryl-CoA reductase
h	Hours
HCMV	Human cytomegalovirus
HSV	Herpes simplex virus
HHV	Human herpes virus
HTLV	Human T-cell Leukaemia virus
hTE	Peroxisomal acyl-CoA thioesterase
HPV	Human papilloma virus
HPI	Hours post-infection
IC50	Inhibitory concentration 50
ICAM-1	Intercellular adhesion molecule 1
IFN	Interferon
IL	Interleukin
IN	Integrase
IMP	Inosine monophosphate
IPTG	Isopropylthio-b-D-galactoside
IRP2	Iron regulatory protein-2
KEGG	Kyoto Encyclopedia of Genes and Genomes

Kb	Kilobase
KS	Kaposi's sarcoma
KSHV	Kaposi's sarcoma-associated herpes virus
LAG-1	Longevity-assurance gene 1
LAV	Lymphadenopathy Associated Virus
LB	Luria-Bertani
LC	Langerhans cell
LDL	Low density lipoprotein
LFA-1	Lymphocyte function-associated antigen 1
Log(2)	Log base 2
LTNP	Long-term non-progressor
LTR	Long terminal repeat
MAPK8	Mitogen-activated protein kinase 8
MHC	Major histocompatibility complex
µg	Microgram
µl	Microlitre
µM	Micromolar
MA	Matirx
MIAME	Minimal Information about a Microarray Experiment
MIP1-a/b	Macrophage inflammatory protein-a/b
ML	Maximum likelihood
MLV	Murine leukaemia virus
MMP-2	Matrix metalloproteinase
MOI	Multiplicity of infection
MondoA	Mlx interactor
MP	Maximum parsimony
MRC	Medical Research Council
mRNA	Messenger RNA
MT1-MMP	Membrane-type-1-matrix-metalloproteinase
MVB	Multivesicular body
MYC	v-myc myelocytomatosis viral oncogene homolog
N-TEF	Negative transcription elongation factors
NC	Nucleocapsid
NCBI	Natioanl Centre for Biotechnology Information
Nedd4	Neural precursor cell expressed, developmentally down-regulated 4
NEK3	NIMA (never in mitosis gene a)-related kinase 3
Nef	Negative factor
NFAT	Nuclear factor of activated T-cells
ng	Nanogram
NJ	Neighbour joining
NRTI	Nucleoside Reverse Transcriptase Inhibitors
NNRTI	Non- Nucleoside Reverse Transcriptase Inhibitors
nSI	Non- syncytium inducing
NXF1	Nuclear RNA export factor 1
OH	Hydroxyl
PE	Phycoerythrin
PBMC	Peripheral blood mononuclear cell
PBS	Primer binding site/phosphate buffered saline
PDCD2	Programmed cell death-2
PCR	Polymerase chain reaction
PHA	Phytohaemagglutinin
PI	Protease inhibitor or post infection
PIC	Pre-integration complex
Pim2	Pim-2 oncogene
pH	Potential of Hydrogen
PMT	Photomultiplier tube
pol	Polymerase
PCP	Pneumocystis carinii pneumonia
pMol	Picomoles or picomolar
PR-RT	Protease and reverse transcriptase

PSSM	Position specific scoring matix
P-TEF	Positive transcription elongation factors
Puro	Puromycin
PVM	Pneumonia virus of mice
R5	CCR5
rev	Regulator of expression of virion
RFLP	Restriction fragment length polymorphism
RRE	Rev response element
RNA	Ribonucleic acid
RPM	Revolutions per minute
RPMI	Roswell Park Memorial Institute
RSV	Respiratory syncitial virus
RT	Reverse transcriptase or room temperature
RTC	Reverse-transcription complex
RTK	Receptor tyrosine kinase
RTPCR	Reverse transcription-polymerase chain reaction
RXRA	Retinoid X receptor alpha
S100A10	S100 calcium binding protein A10
SAM	Significance analysis of microarrays
SERTAD2	SERTA domain containing 2
SD	Standard deviation
SDF-1	Stromal-derived factor 1
SFRS5	Splicing factor arginine/serine-rich
SI	Syncytium inducing
siRNA	Short interfering RNA
SIV	Simian immunodeficiency virus
SIVA	CD27-binding (Siva) protein
SNR	Signal to noise ratio
SOM	Self-organising map
SREPB2	Sterol responsive element binding protein 2
SSCP	Single strand conformation polymorphism
ssDNA	Strong stop DNA
ssRNA	Single stranded RNA
SSPE	Saline sodium phosphate EDTA
ST1-R5	SupT1-R5 T-cells
STAR	Subtype analyser
STAT1	Signal transducer and activator of transcription 1
STD	Sexually transmitted disease
Tar	Trans-activation response element
tat	Transcriptional transactivator
TBE	Tris-borate-ethylenediaminetetraacetic acid
TBS	Tris buffered saline
TFIIH	General transcription factor IIH
TGFβ	Transforming growth factor beta
TH1	T helper-1
TH2	T helper-2
TIA1	TIA1 cytotoxic granule-associated RNA binding protein
TRNA	Total RNA
T-TNA	Transfer RNA
TIMP2	Tissue inhibitor of matrix metalloproteinase 2
Tip30	Tat-interacting protein 30
TNF	Tumour necrosis factor
TR	Thyroid hormone receptor
TRIP	Thyroid hormone receptor interacting protein
Tsg101	Tumour susceptibility protein-101
U	Units
UHR-RNA	Universal human reference RNA
UNAIDS	Joint United Nations Programme on HIV/AIDS
UV	Ultraviolet
V3	Variable region 3

Vam6	Vacuolar protein sorting 39
VAMP	Vesicle-associated membrane proteinv
VEGF	Vascular endothelial growth factor
vif	Viral infectivity factor
vpu	Viral protein U
Vps4	Vaculolar sorting protein 4
WHO	World Health Organization
X4	CXCR4
X-Gal	5-bromo-4-chloro-3-indolyl-b-D-galactoside

Chapter 1.0 Introduction

1.1 History

In June 1981, the Centres for Disease Control (CDC) published a report about the occurrence, without identifiable cause, of a rare lung infection *Pneumocystis carinii* pneumonia (PCP) in five men in Los Angeles (MMWR weekly, 1981). A few days later, following these reports of PCP and other rare life-threatening opportunistic infections, the CDC formed a Task Force on Kaposi's Sarcoma and Opportunistic Infections (KSOI). At this time a number of theories were developed about the possible cause of these opportunistic infections and cancers. Early theories included infection with cytomegalovirus and the use of amyl nitrite or butyl nitrate "poppers" and "immune overload" (Gottlieb *et al.*, 1981, Goedert *et al.*, 1982). Five months later in December 1981, however, when the first cases of PCP were reported in injecting drug users it was clear that this disease not only affected gay men, but other population groups (Masur *et al.*, 1981). At the same time the first case of such an illness was documented in the UK (du Bois *et al.*, 1981).

By 1982 the disease still did not have a name, with different groups referring to it in different ways. The CDC referred to it by reference to the diseases that were occurring, although on some occasions they referred to it as KSOI, the name already given to the CDC task force. Journals chose a different nomenclature, with the Lancet calling it the 'gay compromise syndrome', whilst at least one newspaper referred to it as GRID (gay-related immune deficiency) (Brennan and Durack, 1981, Altman, 1982) and another newspaper described it as 'gay cancer' and 'community-acquired immune dysfunction' (The Washington Blade, 1982).

In June 1982, a breakthrough was made in the reporting of a group of cases amongst gay men in Southern California, suggesting that this disease might be caused by an infectious agent that was sexually transmitted (MMWR weekly, 1982-a). By the beginning of July a total of 452 cases, from 23 American states, had been reported to the CDC (CDC, 1982). By August the disease was being referred to by its new name of Acquired Immunodeficiency Syndrome, or AIDS (Marx, 1982).

In December 1982, a 20-month old child who had received multiple transfusions of blood and blood products died from infections related to AIDS providing further evidence that AIDS was caused by an infectious agent. At the same time the CDC reported the first cases of possible mother to child transmission of AIDS (MMWR Weekly, 1982-b). A report was also published which stated that a disease previously known as "slim", was actually an African form of AIDS (Kamradt *et al.*, 1985). With the dawn of 1983, it became evident that AIDS could also be transmitted heterosexually from men to women.

It was in May 1983 that researchers at the Institut Pasteur in France reported that they had isolated a new virus, which they believed was the cause of AIDS (Barre-Sinoussi *et al.*, 1983). They had detected reverse transcriptase (RT) activity in the culture supernatant of a lymph node biopsy from a patient with lymphadenopathy, a condition already associated with individuals at risk for AIDS. The presence of RT activity was indicative of the presence of a retrovirus, at the time only two of which were known: Human T-cell Leukaemia virus I (HTLV-I, Gallo *et al.*, 1981) and II (Kalyanaraman *et al.*, 1982). Proteins from the infected cell extracts, however, could not be precipitated with antibodies to either virus (Barre-Sinoussi *et al.*, 1983), suggesting it was distinct from both. It was named Lymphadenopathy Associated Virus (LAV). In 1984, similar reports of virus isolation from AIDS patients were made (Popovic *et al.*, 1984). In this case the virus was named HTLV-III, because of apparent similarities in the host cell preference of this new virus to HTLV-I and –II. Further studies demonstrated the presence of LAV-like viruses in the PBMCs of homosexual men with AIDS, and furthermore the presence of high antibody titres against the virus in the otherwise healthy male homosexual control group (Levy *et al.*, 1984). These viruses were named AIDS-related viruses (ARVs). The subsequent cloning and sequencing of LAV (Alizon *et al.*, 1984), HTLV-III (Ratner *et al.*, 1985) and ARV (Luciw *et al.*, 1984), however, revealed that HTLV-III and ARV were 99% and 95% similar to LAV, respectively. This implied that the viruses isolated in separate laboratories were actually the same (Montagnier *et al.*, 2002). Numerous studies where HTLV-III/LAV/ARV was isolated from people either at risk or suffering from AIDS provided further evidence that this virus was causing the disease (Gallo *et al.*, 1984, Casareale *et al.*, 1984, Safai *et al.*, 1984, Cheingsong-Popov *et al.*, 1984, Sarngadharan *et al.*, 1984, Goedert and Gallo, 1985).

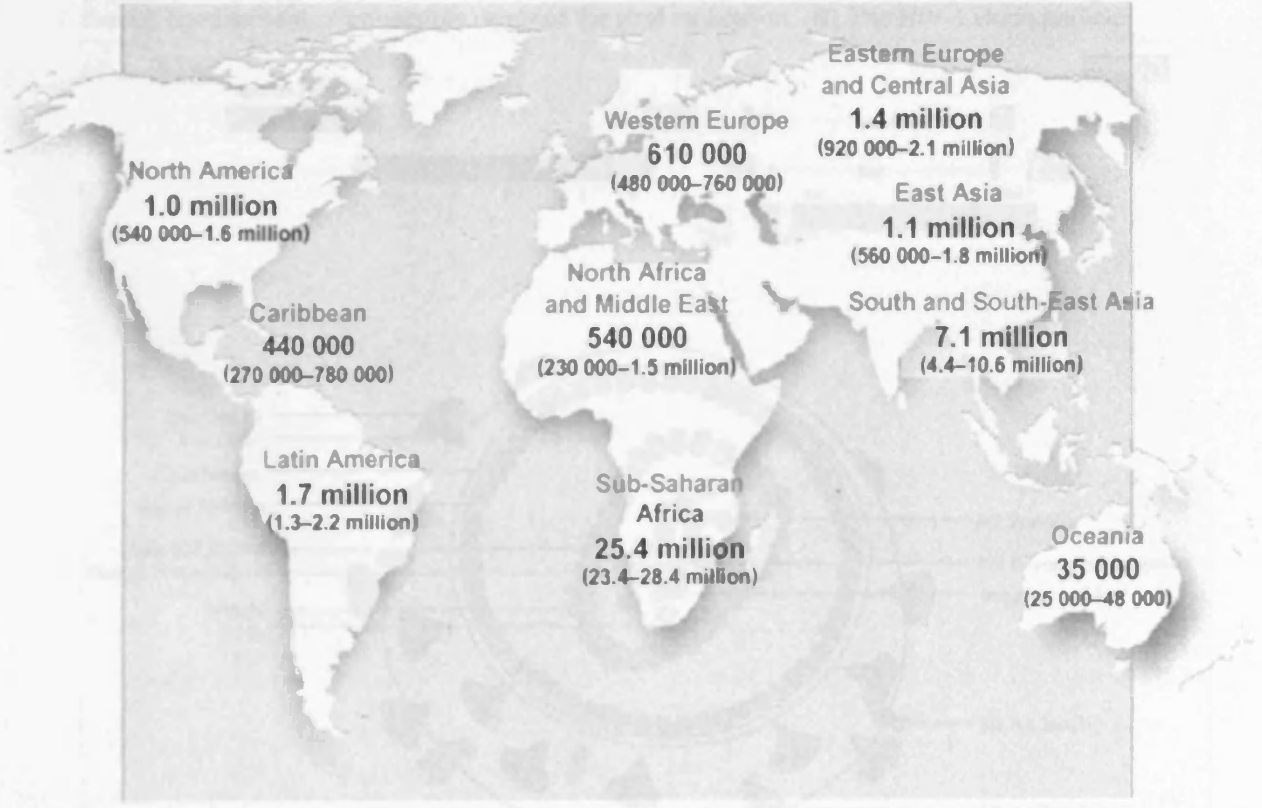
In the meantime, in Europe, two separate AIDS epidemics were occurring. In France and Belgium AIDS was occurring mainly in people from Central Africa or those with links to the area, whilst in the UK, West Germany and Denmark the majority of people with AIDS were homosexual (Weller *et al.*, 1984). In the UK people in risk groups considered particularly susceptible to AIDS were asked not to donate blood (Gunson, 1986). The first meeting to assess the global AIDS situation took place, which initiated global surveillance by the World Health Organisation (WHO). It was reported that AIDS was present in the U.S.A., Canada, fifteen European countries, Haiti, and Zaire as well as in seven Latin American countries. There were also cases reported from Australia and two suspected cases in Japan (WHO, 1983). By the end of 1984, there had been 7699 AIDS cases and 3665 AIDS deaths in the USA, and 762 cases had been reported in Europe (AIDS Activity, CDC, 1984, MMWR Weekly, 1985). In the UK there had been 108 cases and 46 deaths (Department of Health & Social Security, 1985). In 1985, CDC classified routes of transmission as 'intimate sexual contact; sharing contaminated needles; transfusion of whole blood, blood cellular components, plasma or clotting factor concentrates that have not been heat treated; and from an infected mother to her child before, at, or shortly after birth' (MMWR, 1985).

At the same time, controversy regarding the nomenclature of HTLV-III/LAV/ARV came to a head (Marx, 1985). Gallo and colleagues maintained that the virus should be named HTLV-III, because of its predilection for infecting T-cells of the helper class. He also changed the HTLV acronym from meaning Human T-cell Leukaemia virus to Human T-cell Lymphotropic virus to reflect both this, and the fact that HTLV-III was not associated with T-cell proliferation and leukaemias, but rather T-cell death. Montagnier, Levy and others, however, felt that these distinguishing features meant that HTLV-III/LAV/ARV required a new name. Sequencing had shown that, whilst having all the sequence features of a retrovirus, HTLV-III/LAV/ARV also possessed additional open reading frames, more in common with a subgroup of the retroviruses known as the lentiviruses (Wain-Hobson, Alizon and Montagnier, 1985, Sonigo *et al.*, 1985), furthering their cause. Accordingly, a meeting of the International Committee of the Taxonomy of Viruses was convened by Harold Varmus and, amongst others, the main protagonists (Montagnier, Gallo and Levy) came together. A decision was made that the virus should be named Human Immunodeficiency Virus (HIV), later modified to HIV-1 with the discovery of a variant of this virus in West Africa, termed HIV-2 (Clavel *et al.*,

1986-a,b). At the time, however, Varmus was quoted as saying "..these deliberations can be irrelevant to the way people behave, nothing we do is binding. If someone wants to ignore it, he can." Despite this, HIV has passed into common parlance as the virus that causes AIDS.

By the end of the next year, 85 countries had reported 38,401 cases of AIDS to the WHO (Bureau of Hygiene & Tropical Diseases, 1986). Eighteen years later, an estimated 39.4 million people are living with HIV, with 4.9 million new infections and 3.1 million deaths due to AIDS in the past year alone. Some 65% of these infections affect people in sub-Saharan Africa (Figure 1.0).

Figure 1.0 **Adults and children estimated to be living with HIV as of the end of 2004.**

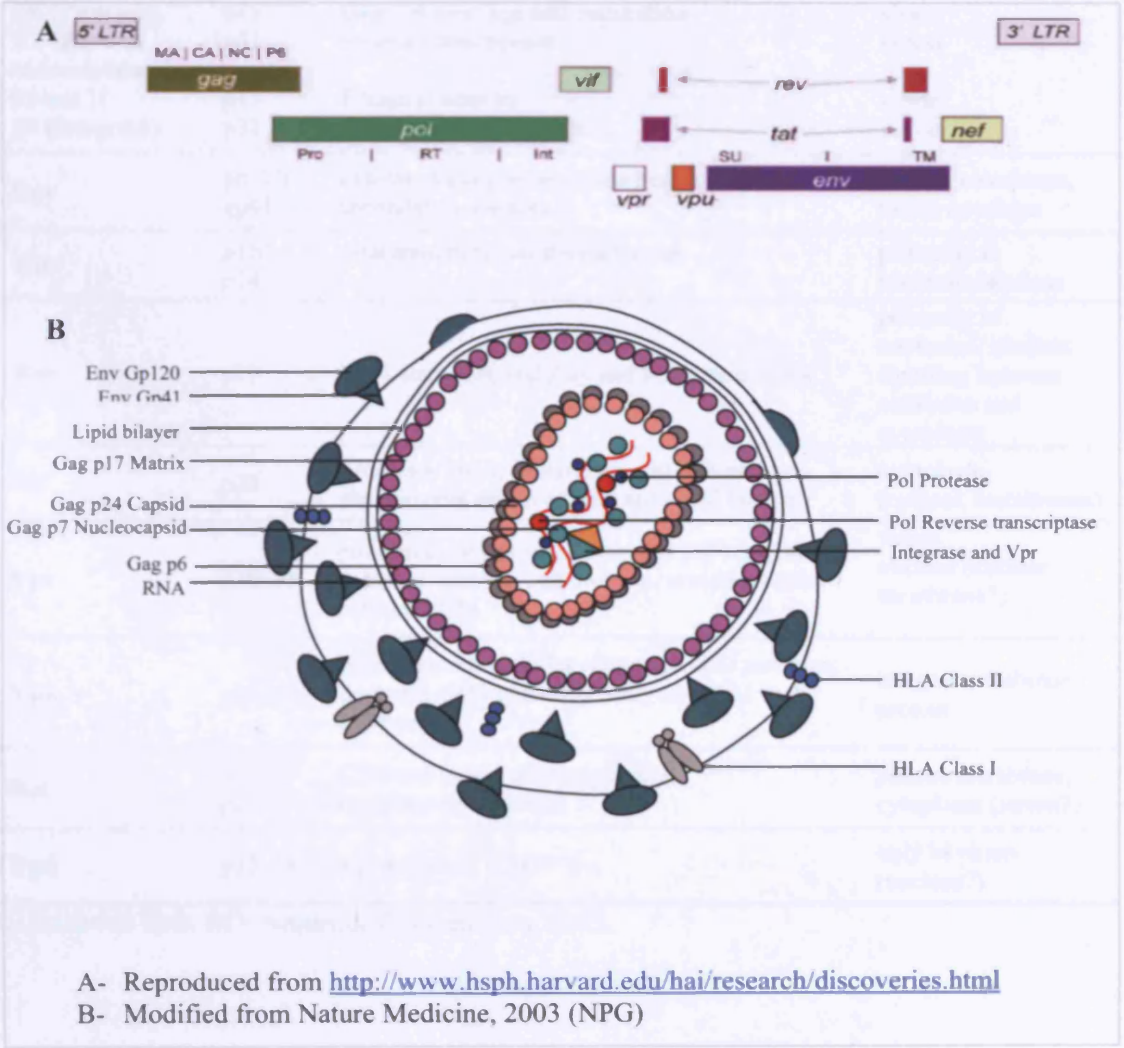


Source: AIDS Epidemic Update, December 2004.

1.2 HIV Structure

HIV-1 and –2 are the only members of the Lentivirus genus known to affect humans, within the family Retroviridae. Whilst retroviruses are characterised by the presence of certain features, including a genome composed of ribonucleic acid (RNA); a common structure encoded by three polyprotein genes: group specific antigen (*gag*), polymerase (*pol*) and envelope (*env*); a life cycle in which the viral genome is reverse transcribed and inserted into the genetic material of the host; and the ability to alter their genomes rapidly by mutation in response to environmental conditions, lentiviruses have additional complex combinations of genes and as the name implies are 'slow' viruses with the potential for long periods of latent infection, prior to causing disease. Figure 1.1 shows the structure of a mature HIV-1 virion and the organisation of its genome.

Figure 1.1 **HIV-1 genome organisation and viral structure.**
(A) Several coding regions overlap and others are discontinuous. Splicing variations generate the full complement of transcripts required for viral replication. **(B)** The HIV-1 virion particle.



A- Reproduced from http://www.hsph.harvard.edu/hai/research/discoveries.html
B- Modified from Nature Medicine, 2003 (NPG)

As depicted in Figure 1.1, the virus is composed of six structural proteins that surround and protect two strands of positive sense genomic RNA and which mediate interactions with host cells. In addition there are three enzymes and six accessory proteins, which function in a variety of processes during viral replication. The size, function and localization of each of these proteins are detailed in Table 1.0.

Table 1.0 Viral protein components of the HIV virion.

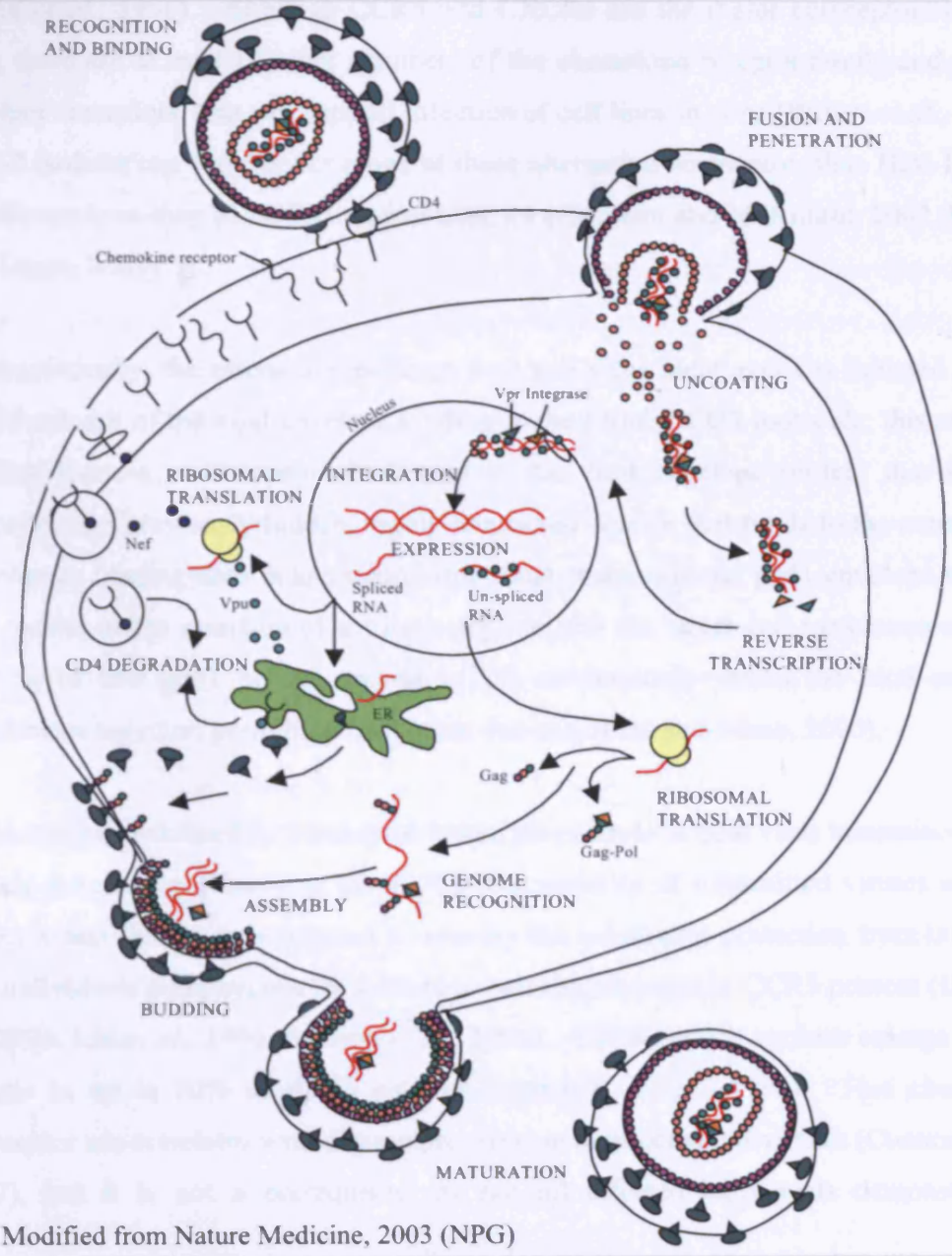
NAME	SIZE	FUNCTION	LOCALISATION
Gag			
MA (matrix)	p17	membrane anchoring; Env interaction; nuclear transport of viral core	virion
CA (capsid)	p24	core structure	virion
NC (nucleocapsid)	p7	binds RNA	virion
p6	p6	binds Vpr	virion
Pol			
PR (Protease)	p15	Gag/Pol cleavage and maturation	virion
RT (Reverse transcriptase)	p51	reverse transcription	virion
RNase H	p15	RNase H activity	virion
IN (Integrase)	p31	DNA provirus integration	virion
Env	gp120/ gp41	external viral glycoproteins bind to CD4 and secondary receptors	plasma membrane, virion envelope
Tat	p16/ p14	viral trancriptional transactivator	primarily in nucleolus/nucleus
Rev	p19	RNA transport, stability and utilization factor	primarily in nucleolus/ nucleus shuttling between nucleolus and cytoplasm
Vif	p23	promotes virion maturation and infectivity - circumvents action of host anti-viral factors	cytoplasm (cytosol, membranes)
Vpr	p10–15	promotes nuclear localisation of pre-integration complex, inhibits cell division, arrests infected cells at G2/M	virion, nucleus (nuclear membrane?)
Vpu	p16	promotes extracellular release of viral particles; degrades CD4 in the ER; only in HIV-1 and SIVcpz	integral membrane protein
Nef	p25 - p27	CD4 and class I downregulation (myristylated protein)	plasma membrane, cytoplasm (virion?)
Vpx	p12 16	Vpr homolog in HIV-2	only in virion (nucleus?)

Reproduced from HIV Sequence Compendium, 2003.

1.3 HIV Replication

The retroviral life cycle is typified by a number of processes. Following entry to a cell these include: reverse transcription of the viral ssRNA genome to dsDNA; incorporation of viral DNA into the chromosomes of the host cell where it is known as a provirus; and transcription of this integrated genome by the host's RNA polymerase enzymes to create new viral ssRNA. This RNA can then function as mRNA to make viral proteins or be packaged as genome into new viral particles. An overview of the HIV replication is illustrated is Figure 1.2.

Figure 1.2 Overview of the HIV replication cycle.



Modified from Nature Medicine, 2003 (NPG)

1.3.1 Host cells, binding and fusion

The main cell surface molecule to which HIV binds with high affinity is CD4 (Dalgleish *et al.*, 1984, Klatzmann *et al.*, 1984). This cell-surface glycoprotein is present at low concentration on monocytes, macrophages and antigen presenting dendritic cells. It is present in abundance, however, on the surface of immature T lymphocytes and mature circulating T-helper cells (Sattentau and Weiss, 1988), making these cells a prime target for infection. In addition to CD4, HIV-1 requires a co-receptor for cell entry, the type of which defines two 'groups' of HIV: macrophage tropic strains, which use the β-chemokine receptor CCR5 (Dragic *et al.*, 1996, Alkhatib *et al.*, 1996); and T-cell tropic strains, which use the α-chemokine receptor CXCR4 (Feng *et al.*, 1996). Although CCR5 and CXCR4 are the major coreceptors used *in vivo*, there are at least 12 other members of the chemokine receptor family and related "orphan" receptors, that can support infection of cell lines *in vitro* (Willey *et al.*, 2003). HIV-2 isolates can use a wider range of these alternative coreceptors than HIV-1, often as efficiently as they use CCR5 and/or CXCR4 (Clapham and McKnight, 2002, Reeves and Doms, 2002).

Mechanistically, the interaction between host and virus membranes is initiated by the gp120 subunit of the viral envelope binding to the cellular CD4 molecule; this receptor binding induces conformational changes in the viral envelope protein that include exposure of a previously hidden, highly conserved domain that binds to the coreceptor. Coreceptor binding then induces conformational changes in the gp41 envelope subunit that results in the insertion of a fusion peptide into the target cell membrane and the binding of two gp41 helical regions. This mechanically draws the viral and cell membranes together, permitting membrane fusion (Doms and Trono, 2000).

Cell tropism, as defined by coreceptor usage, plays a role in both virus transmission and disease progression (Huang *et al.*, 1996). The majority of transmitted viruses are R5-tropic, a fact that is demonstrated *in vivo* by the substantial protection from infection that individuals homozygous for a 32-base pair (bp) deletion in CCR5 possess (Dean *et al.*, 1996, Liu *et al.*, 1996, Samson *et al.*, 1996). CXCR4-using variants emerge late in disease in up to 50% of AIDS patients (Tersmette *et al.*, 1988). This change in coreceptor use correlates with disease progression in infected individuals (Connor *et al.*, 1997), but it is not a prerequisite, as not all infected individuals demonstrate a

coreceptor switch (de Roda Husman *et al.*, 1999). Furthermore, in patients infected with a particular genetic subtype of HIV-1 (C) viral switching to X4-usage is rarely detected. Although primary X4 strains can infect macrophages via CXCR4 (Simmons *et al.*, 1998, Valentin *et al.*, 2000), these variants primarily target new populations of T cells that express CXCR4 but not CCR5, such as naive T-cells (Blaak *et al.*, 2000). Whilst the mechanisms of receptor binding, conformational changes, fusion and the subsequent process of reverse transcription are now fairly well defined, very little is known about the concurrent uncoating process.

1.3.2 Uncoating and reverse transcription

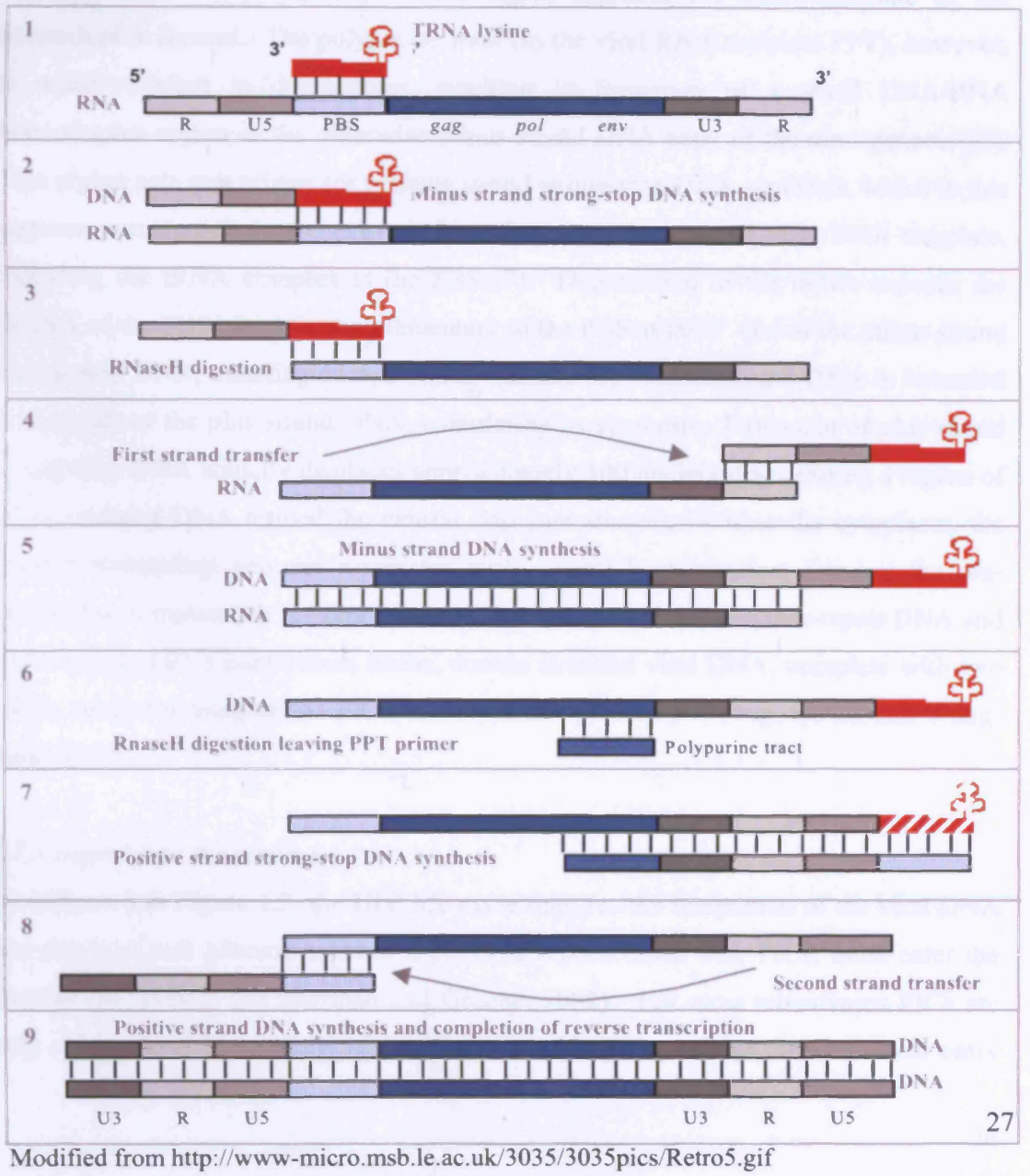
The fusion of viral and cellular membranes delivers the viral core into the cytoplasm, where the viral RNA is reverse transcribed by the virion-packaged reverse transcriptase (RT), generating a linear double-stranded DNA molecule. Although there is evidence for limited DNA synthesis in virions prior to infection (Lori *et al.*, 1992, Trono *et al.*, 1992), reverse transcription usually occurs after the release of the viral core into the cytoplasm of the target cell. The trigger for the initiation of reverse transcription is not clearly understood, but exposure of the incoming viral ribonucleoprotein complex to a significant concentration of deoxyribonucleotides in the cytoplasm is thought to play an important role (reviewed in Goff, 2001).

Specifically, following its release into the cytoplasm, the viral core undergoes a partial and progressive disassembly (uncoating). This results in the generation of subviral particles termed reverse-transcription complexes (RTCs) and pre-integration complexes (PICs). Some viral and cellular proteins appear to influence the uncoating and/or the reverse-transcription of retroviruses, including HIV-1 Nef, Vif and the cellular protein cyclophilin A (CypA), and in some cases their mode of action is not entirely clear (reviewed in Nisole and Ali Saïb, 2004). In the case of Vif, this protein has been shown to counteract the antiviral activity of CEM15/APOBEC3G by preventing its incorporation into progeny virions (Lecossier *et al.*, 2003, Mangeat *et al.*, 2003, Marin *et al.*, 2003, Sheehy *et al.*, 2003). This cellular protein inhibits HIV replication at the step of reverse-transcription, so this finding is consistent with the prior observation that viruses lacking Vif are unable to complete viral DNA synthesis (von Schwedler *et al.*, 1993). In a similar interaction, CypA is believed to protect the viral capsid from the

human restriction factor Ref1, leading to an increase in HIV-1 infectivity (Towers *et al.*, 2003).

Reverse transcription proceeds within RTCs. This process is catalysed by the viral reverse transcriptase and involves the synthesis of several DNA intermediates. Through repositioning of these on the RNA genome template, self-priming enables the synthesis of double stranded DNA. A schematic representation of this process is shown in Figure 1.3. Due to the discontinuous nature of this process, recombination between different viral genomes, within the same cell, may take place (Tang, Kuhen and Wong-Staal, 1999).

Figure 1.3 Schematic of reverse transcription of the HIV genome.



Modified from http://www-micro.msb.le.ac.uk/3035/3035pics/Retro5.gif

As shown in Figure 1.3, the first event in reverse transcription of the HIV genome is synthesis of strong-stop DNA, mediated by the viral reverse transcriptase (RT). This sequence corresponds to the extreme 5' end of the viral RNA genome template and is generated by binding of the 3' terminal 18-19 nucleotides of a host-derived transfer RNA (tRNA) lysine to a complementary sequence in the viral RNA, the primer binding site (PBS) (1). Strong-stop DNA is synthesised (2), the integral RNAse H activity of the RT enzyme subsequently degrading the viral RNA in the RNA-DNA heteroduplex (3). This reveals complementary R regions (within the LTR) at both 5' and 3' ends of the genome, which enables the hybridisation of strong-stop DNA to the 3' end of the RNA template genome, referred to as first strand transfer (4). The strong-stop DNA now acts as a primer for minus strand DNA synthesis, which forms the complete 3' LTR (5). RNAse H activity of RT again degrades the RNA template in the heteroduplex formed. The polypurine tract (in the viral RNA template PPT), however, is more resistant to degradation, resulting in formation of a small DNA/RNA heteroduplex region in the otherwise minus strand DNA copy of the viral genome (6). This region acts as a primer for positive strand strong-stop DNA synthesis, which in this instance corresponds to the extreme 3' end of the minus strand viral DNA template, including the tRNA complex at the PBS (7). Degradation of the tRNA exposes the ssDNA of the PBS which is complementary to the PBS at the 3' end of the minus strand strong-stop DNA, enabling second strand transfer (8). Minus strand DNA is extended to the ends of the plus strand DNA, completing its synthesis. Extension of plus strand strong-stop DNA actually displaces approximately 100 nucleotides, creating a region of triple-stranded DNA termed the central flap (not shown). Within the cytoplasm, the reverse transcribed genome associates with several host proteins, forming the pre-integration complex (PIC). Once within the nucleus, cellular enzymes repair DNA and nicks to generate a continuous, linear, double stranded viral DNA, complete with two LTRs, ready for integration. For a review of this process see Tang, Kuhen and Wong-Staal, 1999.

1.3.3 Import into the nucleus

As indicated in Figure 1.2, the HIV life cycle requires the integration of the viral DNA into the host cell genome to form a provirus. To achieve this, PICs, must enter the nucleus (for review see Sherman and Greene, 2002). For most retroviruses PICs are only able to enter the nucleus when the cell is undergoing mitosis, hence nuclear entry

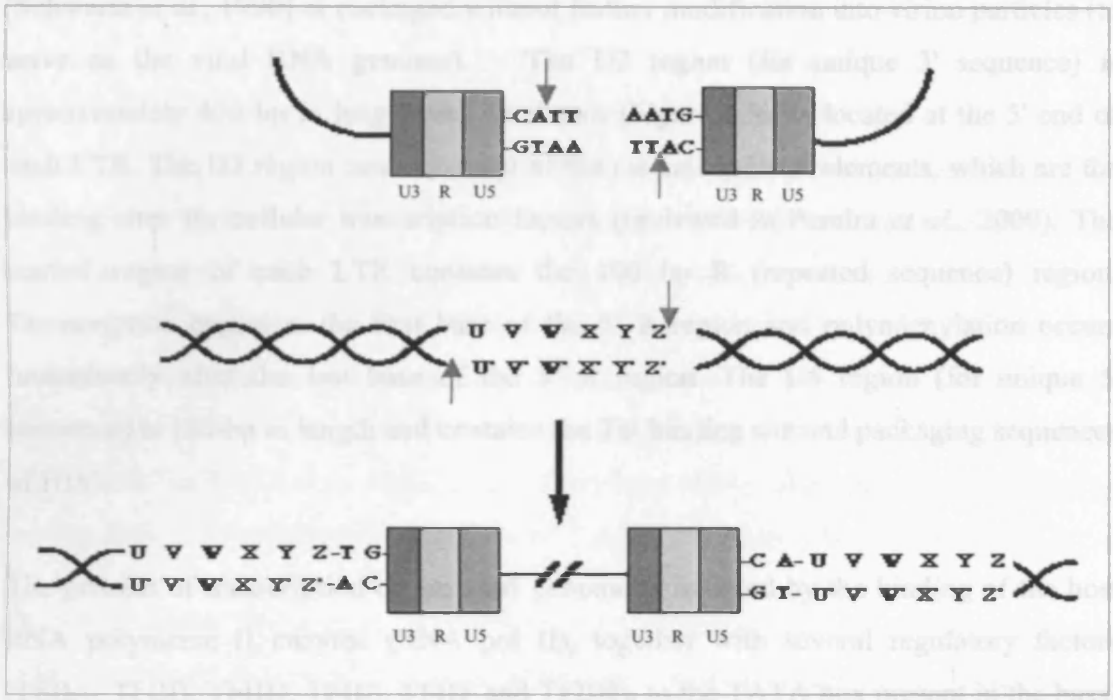
is dictated by the cell cycle and replication is not possible within non-dividing cells. Lentiviruses such as HIV, however, can productively infect non-dividing cells such as macrophages or quiescent T lymphocytes, indicating that PICs are able to actively cross the nuclear membrane (Weinberg *et al.*, 1991, Bukrinsky *et al.*, 1992). Since HIV PICs, composed of the double-stranded linear DNA associated with the viral proteins MA, RT, IN and Vpr have an estimated diameter of 56 nm (Miller, Farnet and Bushman, 1997) and the central channel of the nuclear pore has a maximum diameter of 25 nm, HIV has developed a strategy to enable passage through these structures. The actual mechanism, however, remains elusive.

It has been proposed that HIV uses the host cell cytosolic nuclear transport machinery to transfer its DNA into the nucleus, via nuclear pores, based on the presence of nuclear localisation signals in components of the PIC including the IN and MA proteins (Bukrinsky *et al.*, 1992). Involvement of Vpr (another protein component of PICs) in nuclear import, however, is one of the most controversial. It has been found to interact directly with components of the nuclear pore complex, which is believed to enhance nuclear import efficiency (Popov *et al.*, 1998 a,b). In addition, Vpr expression has been shown to induce transient bulges in the nuclear envelope, which can rupture, creating a channel between the nucleus and the cytoplasm through which PICs may pass (de Noronha *et al.*, 2001). The precise role of these nuclear envelope disruptions, however, remains uncertain as Vpr-deficient viruses can infect non-dividing cells efficiently (Bouyac-Bertoia *et al.*, 2001, Reil *et al.*, 1998). In contrast, the Vpx protein encoded by HIV-2 and SIV has been shown to be both necessary and sufficient for the nuclear import of PICs (Fletcher *et al.*, 1996). Finally, a non-protein component of the PIC that has been described to be important for nuclear entry is the central flap, created during reverse transcription of the viral RNA genome. It has been proposed to act as a cis-acting determinant in nuclear import as its elimination has been shown to result in accumulation of linear viral genomic DNA in the cytosol (Zennou *et al.*, 2000).

1.3.4 Integration into the host genome

Once within the host cell nucleus, integration of viral DNA into the host cell genome proceeds. Figure 1.4 shows a schematic of the integration process.

Figure 1.4 Schematic of viral integration.



Modified from http://www-micro.msb.le.ac.uk/3035/3035pics/Retro7.gif

The process of integration is initiated and catalysed by the association of the viral protein IN with each end of the linear double stranded viral DNA genome. Here, it removes two nucleotides from the 3' terminus of each strand, and catalyses a nucleophilic attack of the recessed 3' hydroxyl (OH) groups on phosphodiester bonds in each of the target cellular DNA strands. Each strand of the viral DNA is then joined to the cellular DNA, leaving a 4-6 bp mismatch at each end. Cellular gap repair enzymes repair these, leaving a 4-6 bp duplication at both the 5' and the 3' junction between host cellular DNA and viral DNA (Hindmarsh and Leis, 1999). The integration process is believed to be semi-random, in that there are certain 'hot-spots' where integration occurs with greater frequency, which typically correspond to transcriptionally more active sites within host chromosomes (Schroder *et al.*, 2002).

1.3.5 Transcription of viral genome

HIV transcription is mediated by a single promoter in the 5' LTR. Expression from the 5' LTR generates a 9.3 kilobase (kb) primary transcript which is approximately 600 bases shorter than the provirus at the 5' end, and which encodes all nine HIV genes.

This primary transcript can be spliced into one of more than 30 mRNA species (Schwartz *et al.*, 1990) or packaged without further modification into virion particles (to serve as the viral RNA genome). The U3 region (for unique 3' sequence) is approximately 450-bp in length and, as shown (Figure 1.3), is located at the 5' end of each LTR. The U3 region contains most of the cis-acting DNA elements, which are the binding sites for cellular transcription factors (reviewed in Pereira *et al.*, 2000). The central region of each LTR contains the 100 bp R (repeated sequence) region. Transcription begins at the first base of the 5' R region and polyadenylation occurs immediately after the last base of the 3' R region. The U5 region (for unique 5' sequence) is 180-bp in length and contains the Tat binding site and packaging sequences of HIV.

The process of transcription of the viral genome is initiated by the binding of the host RNA polymerse II enzyme (RNA pol II), together with several regulatory factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH), to the TATA box present in the basal promoter region of the 5' LTR U3. The helicase action of TFIIH results in separation of an 11 bp stretch of the DNA and phoshorylation of the C-terminal domain (CTD) of the RNA pol II enzyme, initiating elongation of the transcript. The phosphorylation of the RNA pol II enzyme is regulated by both negative transcription elongation factors (N-TEFs) and positive transcription elongation factors (P-TEFs), the main positive elongation factor involved being P-TEFb, which is composed of two subunits: cyclin T1 and cyclin dependent kinase 9 (cdk9). Specifically, this phosphorylates the CTD of RNA pol II, complexed to the HIV promoter, causing the dissociation of two N-TEFs, DSIF and NELF (Garber and Jones, 1999). In addition to these host factors, however, in order for full-length transcription of the viral genome to occur, the viral protein Tat is required. This protein binds to a short stem-loop structure present in the R-region of each viral transcript, named the trans-activation response element (TAR). This binding facilitates Tat interaction with the cyclin T1 component of PTEF-b, in effect recruiting it into the transcriptional complex and enabling it to phosphorylate the CTD of RNA pol II, resulting in the enhanced transcript elongation required (Daelemans, De Clercq and Vandamme, 2001). In addition, it should be noted that early expression of viral genes from unintegrated viral cDNA has also been described (Butera *et al.*, 1991, Stevenson *et al.*, 1990, Teo *et al.*, 1997, Cara *et al.*, 1996, Wu and Marsh, 2003). Although the role of

this early expression is not clear, it is enhanced in the presence of Vpr (Poon and Chen, 2003).

1.3.6 Nuclear export of viral RNA

The primary HIV-1 transcript contains multiple splice donors (5' splice sites) and splice acceptors (3' splice sites), which can be processed to yield more than 30 alternative mRNAs (Schwartz *et al.*, 1990). HIV-1 mRNAs fall into three size classes: an unspliced 9-kb primary transcript which can be expressed to generate the Gag and Gag-Pol precursor proteins or be packaged into virions to serve as genomic RNA; incompletely spliced RNA which can potentially express Env, Vif, Vpu, Vpr and the single-exon form of Tat (~ 4 kb); and fully spliced RNA, which have the potential to express Rev, Nef, and the two-exon form of Tat (~2 kb) (Malim and Cullen, 1993).

Normally cellular RNAs must be completely spliced before they can exit the nucleus in order to prevent the translation of intronic sequences contained in partially spliced mRNAs. In the case of HIV, therefore, fully spliced viral mRNAs exit the nucleus by using the export pathway followed by the majority of cellular mRNAs, resulting in the immediate expression of Nef, Tat, and Rev. In contrast, the two RNA species that contain intronic sequence (unspliced and singly spliced viral RNAs) require assistance in order to bypass the normal "check point" of RNA splicing and be exported from the nucleus. This is achieved by the action of the regulator of expression of virion (Rev) protein, which binds to viral RNAs that retain intron sequences. This interaction is mediated by a 351 nt region of extensive secondary structure, within the Env coding region, called the Rev response element (RRE) (Mann *et al.*, 1994). Rev contains both nuclear export and import signals, so it may shuttle both in and out of the nucleus, depositing intron containing viral RNAs in the cytosol (Meyer and Malim, 1994). As threshold levels of Rev are necessary for exporting intron-containing HIV mRNAs, this explains why these encode the viral late gene products.

1.3.7 Translation of viral RNA

HIV-1 mRNA, once within the cytosol, is translated into protein by the host cell translation machinery. Structural proteins encoded by the Gag gene are translated from an unspliced, genomic viral RNA as a 55 kDa fusion protein including MA, CA, NC and p6. In approximately one in every twenty Gag translation events, however, the

cellular translation machinery undergoes a –1 frameshift in the p6 coding region. In this event, translation does not stop at the Gag stop codon, but continues on into the adjacent Pol sequence to generate a 160 kDa polyprotein fusion of Gag and Pol, including PR, RT and IN. It is the activity of PR that releases the individual enzymatic protein components within this polyprotein (Cimarelli and Darlix, 2002). A single spliced viral RNA encoding the envelope proteins gp41 and gp120 is targeted to the endoplasmic reticulum (ER), where a gp160 precursor is synthesised alongside other integral membrane proteins synthesised by the host. Here, gp160 is glycosylated and associates into trimers before being cleaved to form gp120 and gp41 within the golgi apparatus. This mature envelope protein is then trafficked to the cell membrane, for viral assembly and budding.

1.3.8 Virion assembly and budding

There are still many processes not fully understood in terms of the location in the host cell at which retrovirus assembly takes place, and the mechanism by which Gag proteins target the subcellular site of assembly. It is widely held, however, that the first stage in virion assembly is the targeting of the Gag precursor to the host cell membrane. This is mediated by the M-domain, at the N-terminus of this polyprotein, which inserts into the plasma membrane and enables electrostatic interaction between the polar phopholipid head groups within the membrane and a number of basic amino acid residues within the M-domain. Each virion requires the incorporation of approximately 1200 Gag molecules, association between which are mediated by interaction domains within NC, MA and CA proteins. The NC protein, within Gag, also includes two Cys-His boxes that bind to the virion packaging signal sequence, present in viral genomic RNA. This ensures the inclusion of viral genome within particles, as they form at the cell surface (Kaplan, 2002).

The gradual accumulation of Gag molecules at the cell membrane results in the formation of 'blebs', which increase in size and curvature until the membrane can fuse behind the viral particle, releasing an immature virion (Gottlinger, 2001). Several host proteins have been implicated in this budding process, including tumour susceptibility protein-101 (Tsg101), ALG-2-interacting protein 1 or ALG-2-interacting protein X (AIP1/Alix) and vacuolar sorting protein-4 (Vps4). In the absence of Tsg101 and Vps4, aberrant viral budding occurs: viral particles remain tethered to the cell membrane by

stalk-like projections (Garrus *et al.*, 2001). Finally, viral particles are only fully infectious once they have matured. This process involves cleavage of Gag into its constituent parts by PR, as described (section 1.3.7): MA lines the inside of the viral envelope, CA monomers condense to form a cylindrical core, which surrounds the NC-coated viral genomic RNA (Briggs *et al.*, 2003, Figure 1.1/1.2).

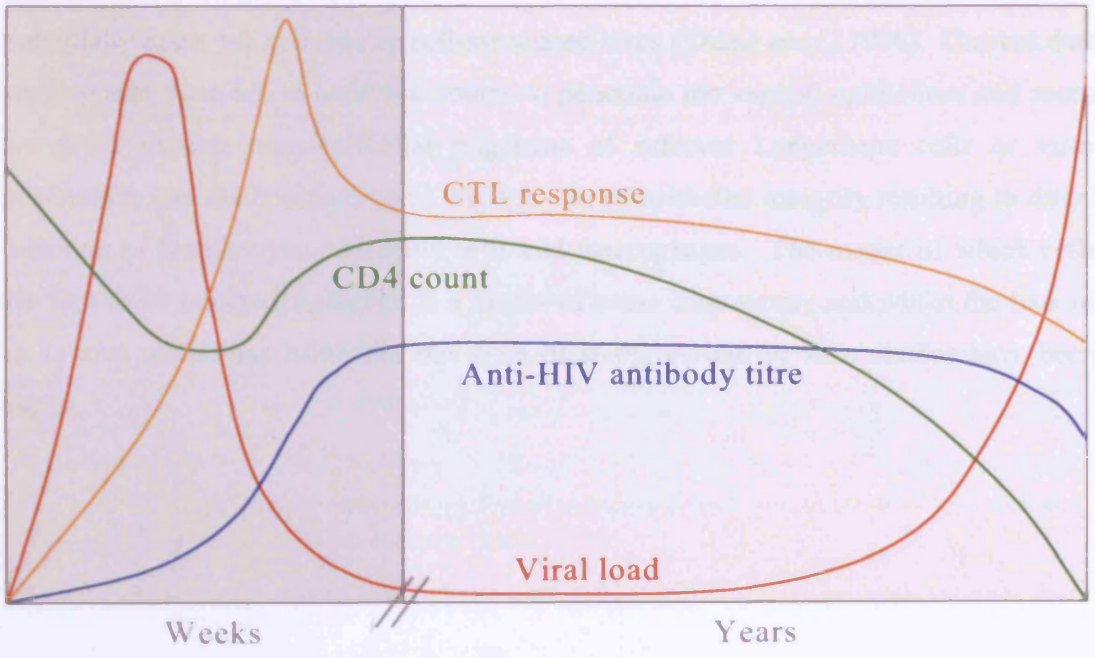
It has been recognized, however, that in certain cell types such as macrophages assembly and budding of HIV virions can take place at an intracellular compartment. Several groups have recently demonstrated that this compartment is the late endosome or multivesicular body (MVB), and it now appears that while the plasma membrane likely represents the predominant site of assembly for these viruses, Gag can also target and assemble in an endosomal compartment in a variety of cell types (reviewed in Freed and Ross, 2004).

In addition, as it buds through the host membrane the HIV virion acquires several host molecules, both within its viral envelope and the particle itself (reviewed in Ott, 2002). There is evidence, however, that this process is not random (Nguyen and Hildreth, 2000). Rather, it is suggested that the virions preferentially assemble and bud through microdomains within the cell membrane termed 'lipid rafts'. Specifically, it has been suggested that the Env glycoproteins traffic with cholesterol and sphingolipids from the golgi apparatus to these rafts (Campbell, Crowe and Mak, 2001). There is evidence that Gag too is targeted to these raft domains, after binding to the membrane (Ono and Freed, 2001). This, in combination with a similarity between the molecular components of lipid rafts and the HIV viral envelope, is suggestive that viral budding and perhaps assembly, occurs at membrane raft domains.

1.4 HIV and AIDS

As described in section 1.1, HIV has been characterised as the causative agent of AIDS, which as a disease may be divided into both acute and chronic phases. The acute phase, in most individuals, is associated with a brief period of flu-like illness (Cooper *et al.*, 1985) and is characterised by a high level of viraemia, the rapid evolution of a cytotoxic T-cell (CTL) response, the synthesis of antibodies directed against the virus and virally infected cells, and a rapid decline in CD4+ T-cells. Despite powerful cell-mediated and humoral responses to HIV infection, however, sterilising immunity is not achieved and a chronic infection is established. Over the course of several years, continuous HIV replication results in CD4+ T-cell attrition, increasing viral load and the occurrence of AIDS-defining infections and cancers, directly due to immune deficiency (Figure 1.5). In the majority of individuals, this chronic phase lasts for approximately eight to ten years. There is, however, some variation in this in that a smaller percentage (approximately 10%) of individuals progress rapidly and develop AIDS within two or three years (rapid progressors). Conversely, an even smaller percentage of HIV infected individuals are termed long-term non-progressors (LTNP) (Sheppard *et al.*, 1993).

Figure 1.5 **Generalised immunological profile of a HIV-1 infected individual with progressive infection.**



By definition, LTNP remain asymptomatic, with normal CD4 cell counts and low or undetectable viral load, in the absence of any antiretroviral therapy. The cause of the lack of progression in these persons is unclear, but it seems to result from the interaction between multiple factors linked to the virus and the host (Anastassopoulou and Kostrikis, 2003). HIV-2 infection is distinct from HIV-1, however, in that the majority of HIV-2 infected individuals display disease progression similar to that of HIV-1 LTNP (Kanki, 1999). The recognition of what factors are the main determinants for protection against disease progression in HIV-1 LTNP and HIV-2 infection are of great interest, since it may allow new treatment strategies to be designed (Rodés *et al.*, 2004).

1.4.1 Transmission and targets for HIV infection

The main routes of HIV-1 transmission are listed in Table 1.1. World-wide, the vast majority of new infections are acquired through heterosexual contact, although in different geographical locations the relative contribution of different transmission routes may differ (Hansasuta and Rowland-Jones, 2001).

Despite there being a wealth of information regarding routes of HIV infection, however, very little is known about the actual molecular processes involved in sexual transmission. In the case of male to female transmission, as semen has been shown to contain both free infectious HIV-1 virions and HIV-1-infected cells, transmission could potentially occur via cell-free or cell-associated virus (Zhang *et al.*, 1998). Current data suggest that there are at least two routes to penetrate the vaginal epithelium and reach lymphoid tissues: trans-epithelial migration of infected Langerhans cells or virus penetration into the lamina propria through loss of epithelial integrity resulting in direct infection of lymphocytes, dendritic cells and macrophages. The matter of which cells are first to be infected, however, is a matter of some controversy and whilst the lack of an *in vivo* model has hampered this field of study, certain *in vitro* studies have been useful.

Table 1.1 Routes and likelihood of HIV-1 transmission.

STUDY POPULATION	TRANSMISSION ROUTE	TRANSMISSION PROBABILITY
Heterosexual	US (male to female)	0.0008–0.001
	US (from male to female, and from female to male)	0.001
	Europe (from male to female, and from female to male)	0.0005–0.001
	Thailand (from male to female)	0.002
	Thailand (from female sex workers to men)	0.03–0.06
	Kenya	0.1
Homosexual	US (receptive anal sex with ejaculation)	0.005–0.03
Perinatal transmission	Thailand	
	- Without AZT	18.9%
	- With AZT	9.4%
	Cote d'Ivoire	
	- Without AZT	24.9%
	- With AZT	15.7%
	USA	
	With Caesarean section and anti-retrovirals:	
	- All three trimesters	2%
	- One or two trimesters	8.2%
	Without Caesarean section but with anti-retrovirals:	
	- All three trimesters	7.3%
	- One or two trimesters	16.4%
	Without caesarean or antiretrovirals	19.0%
Newborn	Breastfeeding:	
	- 6 weeks	3.9%
	- 14 weeks	10.2%
	- 6 months	11.3%
	- 12 months	14.1%
	- 24 months	16.2%
Healthcare workers	Percutaneous (Thailand)	< 0.5%
	Percutaneous (Holland)	0.3%
	Percutaneous (Tanzania)	
	Healthcare workers	0.27%
	Surgeons	0.7%
Blood recipients	Blood transfusion – screened blood	
	- Developed countries	0.003–0.0007%
	- Developing countries	5–10%
Intravenous drug users	Blood contamination	No reports
Artificial insemination with contaminated sperm		Case report

Reproduced from Hansasuta and Rowland-Jones, 2001.

An organ culture of human cervical tissue identified the earliest infected cells to be CD4+ T-cells residing in the lamina propria of the mucosa (Gupta *et al.*, 2002). Whilst rare in the periphery, these predominantly CCR5+/CXCR4-/CD4+ memory T-cells are abundant in mucosal tissues, representing a large proportion of the total T-cell pool, potentially providing an important site for viral replication during acute infection. In terms of Langerhans cells (LCs), these reside in the cervicovaginal epithelium and their normal trafficking pattern to T cell zones of lymph nodes suggests a significant role in HIV transmission (Romani *et al.*, 2001). HIV may come into direct contact with these cells either through exposure of cellular processes to virions in the vaginal lumen or through trauma, which may increase the number of LCs exposed to virus. It is unclear whether exposure to HIV virions alone is sufficient to induce LC migration but skin-derived LCs are susceptible to HIV infection (Kawamura *et al.*, 2000, Tchou *et al.*, 2001) and infected LCs are readily detectable in the oral (Chou *et al.*, 2000) and vaginal mucosa (Bhoopat *et al.*, 2001) of HIV positive individuals. HIV-infected mucosal LCs may therefore mediate transmission of HIV across the intact genital epithelium. In addition to epithelial LCs, CD1a–, CD4+ sub-epithelial dendritic cells (DCs) have a similar capacity for uptake and transport of virus to draining lymph nodes. These cells express C-type lectin receptors including DC-specific intercellular adhesion molecule (ICAM)-grabbing non-integrin (DC-SIGN) (Geijtenbeek *et al.*, 2000), which binds HIV via interaction with high mannose residues expressed on viral gp120. Virus is internalised by the DC and retained in an infectious form for up to 5 days (Pohlmann, Baribaud and Doms, 2001), time enough for the dissemination of infectious virions to the T-cell-rich environment of the lymph nodes without direct DC infection. Infected DCs residing in lymphoid tissue are thus thought to represent an important viral reservoir, throughout the course of disease. Finally, although the prevalence of HIV-1-infected macrophages in mucosal surfaces is low (0.06% of lamina propria mononuclear cells), the extraordinary size of mucosal surfaces (in particular GALT) imparts to intestinal macrophages a prominent role as a HIV-1 reservoir (Smith *et al.*, 2003). Furthermore, monocytes/macrophages may act as a Trojan horse to transmit HIV-1 to the central nervous system (Verani, Gras and Pancino, 2005).

1.4.2 Immune response to HIV infection and immune deficiency

As indicated in Figure 1.5, antibodies can be detected in the blood at high titre during the acute phase of infection. The evolution of a CD8+ T cell-mediated immune

response at this time correlates with the reduction in viraemia shown; antibodies generated at this stage often directed against the HIV-1 structural proteins (Pantaleo and Fauci, 1996). In this acute phase, in addition to viral cytopathic effects, CD8-mediated killing is thought to contribute to CD4+ depletion. It is when the acute phase viraemia is brought under control that CD4+ count can rebound and chronicity is established. During this phase, chronic activation of the immune system can strain homeostasis of naïve and resting memory T cell pools indirectly in a number of ways and, when coupled to the impact of ongoing, low-level destructive events mediated by virus, this strain leads to the progressive depletion of $CD4^+$ T cell pools. Cumulative HIV–induced damage to the thymus and bone marrow, together with age-related reduction in thymic output, also limits the capacity of the body to replenish the naïve T-cell pool during chronic infection, contributing to the gradual CD4+ and CD8+ T-cell decline. Maintenance of a strong cell-mediated response to HIV-1 infection has been associated with non-progressive infection (Copeland and Heeney, 1996), but viral escape from immune control is a defining feature of HIV infection. The ability of HIV to rapidly mutate and evolve under immune pressure ultimately renders both CTL and antibody responses sub-optimal, and viral replication cannot be contained indefinitely. It is also the case that HIV has evolved specific and directed mechanisms to interfere with host immune responses, including down regulation of MHC class I on infected cells by the accessory protein Nef.

It is, however, not only loss of T-cells that results in the development of immune deficiency. Polyclonal B-cell activation and production of autoantibodies contributes to a widespread immune dysfunction, alongside a reduction in natural killer (NK) cell and cytotoxic T-cell activity. T-helper cells also become less responsive to antigens, possibly due to: gp120 and Tat-induced anergy (Copeland and Heeney, 1996); defects in the function of antigen presenting cells (APCs) (Donaghy *et al.*, 2003); and perturbations in the cytokine signalling network (Shearer and Clerici, 1998). It seems that destruction of the immune system is generally not part of HIV's 'strategy', but rather AIDS is a consequence of incomplete adaptation of this virus to a relatively new host. For a concise review of T-cell dynamics in HIV-1 infection see Douek, Picker and Koup, 2003.

1.4.3 Features of HIV disease and therapy

As described in section 1.1, HIV infection was first recognised due to a specific complex of illnesses in otherwise healthy individuals, characteristic of immune deficiency. Today, the complex of illnesses that are associated with and indeed define AIDS, is very well characterised. The WHO disease staging system is shown in Table 1.2. Early in the epidemic, however, treatment of AIDS was in fact treatment of these opportunistic infections resulting from immune deficiency. It was not until 1989 and development of the first directed anti-HIV drug that treatment of the viral cause of immune deficiency began. ACTG019 was a trial of the drug zidovudine, or AZT, which showed that AZT could slow progression to AIDS in HIV positive individuals in the absence of symptoms. In the same year the second drug for the treatment of AIDS, dideoxyinosine (ddI), was made available to people with AIDS, even though only preliminary tests had been completed. Today, there are currently 20 approved antiretroviral drugs in the UK and many more in the expanded access programmes and trials. In terms of classification these anti-virals can be split into four main groups: nucleoside reverse transcriptase inhibitors (NRTIs); non-nucleoside reverse transcriptase inhibitors (NNRTIs); protease inhibitors (PIs); and fusion or entry inhibitors. These are detailed comprehensively in Stolk and Lüers, 2004, and at http://www.avert.org/introtrt.htm. In current times the use of three or more drugs in combination, termed highly active antiretroviral therapy (HAART), has greatly improved the prognosis of HIV positive individuals by effective suppression of viral replication for long periods of time.

The hypermutability of the HIV genome due to the lack of proofreading ability of the viral reverse transcriptase (see section 1.5.2.1), however, means that drug resistant viruses rapidly evolve under drug selection pressure. Experience has shown that by reducing viral load to below the limit of detection on initiation of HAART, this risk of developing drug resistance can be greatly reduced. In the longer term this may reduce the deterioration in immune function and risk of clinical disease progression.

Table 1.2 WHO disease staging system for HIV Infection and Disease in Adults and Adolescents.

CLINICAL STAGE I:
Asymptomatic
Generalized lymphadenopathy
Performance scale 1: asymptomatic, normal activity
CLINICAL STAGE II:
Weight loss, < 10% of body weight
Minor mucocutaneous manifestations (seborrheic dermatitis, prurigo, fungal nail infections, recurrent oral ulcerations, angular cheilitis)
Herpes zoster within the last five years
Recurrent upper respiratory tract infections (i.e. bacterial sinusitis)
Performance scale 2: symptomatic, normal activity
CLINICAL STAGE III:
Weight loss, > 10% of body weight
Unexplained chronic diarrhoea > 1 month
Unexplained prolonged fever (intermittent or constant), > 1 month
Oral candidiasis (thrush)
Oral hairy leucoplakia
Pulmonary tuberculosis
Severe bacterial infections (i.e. pneumonia, pyomyositis)
Performance scale 3: bedridden < 50% of the day during last month
CLINICAL STAGE IV:
HIV wasting syndrome
Pneumocystis carinii pneumonia
Toxoplasmosis of the brain
Cryptosporidiosis with diarrhoea > 1 month
Cryptococcosis, extrapulmonary
Cytomegalovirus disease of an organ other than liver, spleen or lymph node (e.g. retinitis)
Herpes simplex virus infection, mucocutaneous (>1 month) or visceral
Progressive multifocal leucoencephalopathy
Any disseminated endemic mycosis
Candidiasis of esophagus, trachea, bronchi
Atypical mycobacteriosis, disseminated or lungs
Non-typhoid Salmonella septicemia
Extrapulmonary tuberculosis
Lymphoma
Kaposi's sarcoma
HIV encephalopathy
Performance scale 4: bedridden > 50% of the day during last month

Source: http://www.avert.org/hivstages.htm

Compared to monotherapy, therefore, the use of HAART is preferable, but ultimately patients on HAART do fail therapy in time. Estimates are that 1/10 AIDS patients in the UK have failed HAART and require 'salvage therapy', and 1/50 AIDS patients harbour virus that is resistant to all drugs currently available (Kenyon, 2001). Given an increased understanding of this cost/benefit relationship for patients on HAART, there

has recently been a move towards more conservative therapy, due to concerns about relative patient benefit versus toxicities and the difficulties of adhering to a complicated regimen, in the long term (Mocroft and Lundgren, 2004).

1.5 HIV Genetic Diversity and Nomenclature

One of the greatest influences in our understanding of the origin of HIV was the discovery that closely related viruses were present in a wide variety of African primates. HIV and these viruses, the simian immunodeficiency viruses (SIVs), together comprise the primate lentiviruses. With the exception of laboratory-associated infections of Asian macaques, however, studies of the natural history of SIVs in wild primate populations suggest that these close relatives of HIV do not cause disease in their hosts. The implications of this are crucial in our understanding of HIV evolution and pathogenesis.

1.5.1 HIV nomenclature

Following publication of the first complete HIV-1 genome sequence in 1985 (Ratner *et al.*, 1985), extensive sequencing revealed a wide range of genetic variants of this virus infecting people worldwide. First attempts to classify HIV-1 sequences were made in 1988, the approach being to separate viruses into European/North American and African strains, based on the divergence of these sequence groups in phylogenetic trees (Li *et al.* 1988; Smith *et al.* 1988). As additional sequences from other geographic regions were generated, however, it became apparent that this method of classification was too simple. Instead, phylogenetic analysis of *env* gene sequences revealed multiple phylogenetic clusters (referred to as 'clades' or 'subtypes'). Today, HIV-1 has formally been split into three lineages; M (major), O (outlier) and N (non-M/O) (Simon *et al.*, 1998). These are inferred to correspond to separate zoonotic transmission events of simian immunodeficiency virus (SIV-cpz) from the chimpanzee to humans (Figure 1.6). Group M accounts for the majority of HIV-1 infections worldwide and has been classified into 9 major subtypes termed A-D, F-H, J and K.

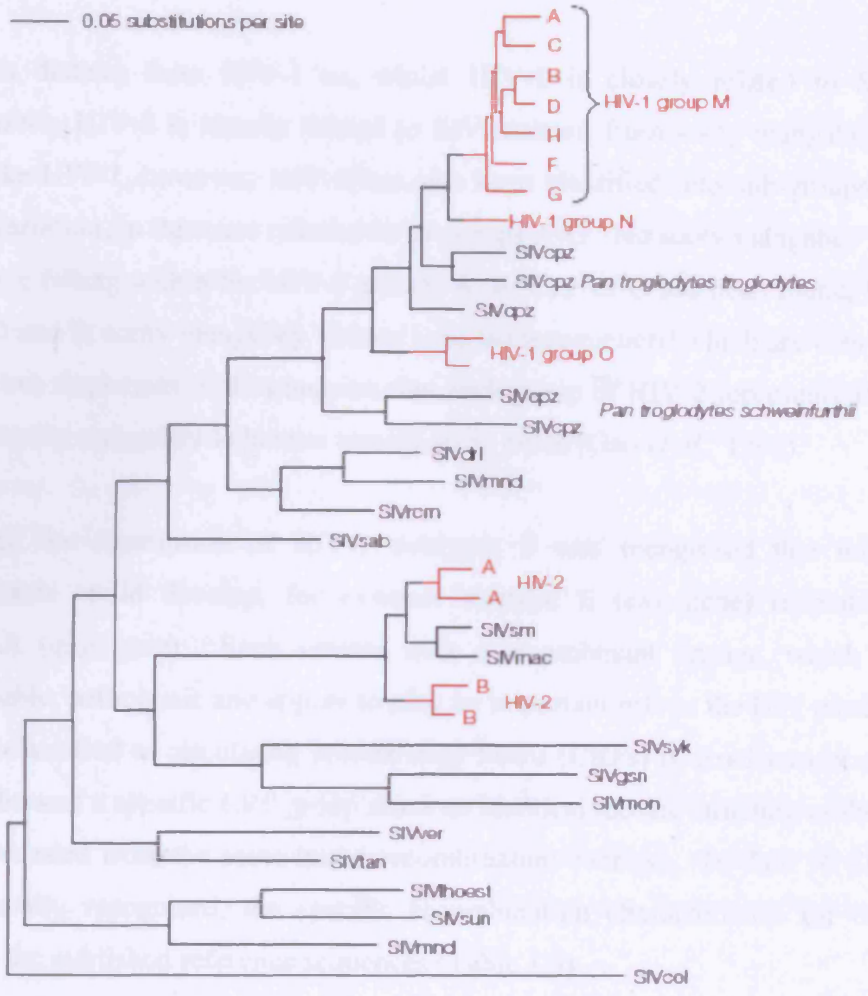


Figure 1.6 Tree showing the evolutionary history of the primate lentiviruses. Because both the HIV-1 and HIV-2 lineages (red branches) fall within the simian immunodeficiency viruses (SIVs) that are isolated from other primates, they represent independent cross-species transmission events. The tree and other evidence also indicate that HIV-1 groups M, N and O represent separate transfers from chimpanzees (SIVcpz), because there is a mixing of the HIV-1 and SIV lineages. Similarly, HIV-2 seems to have been transferred from sooty mangabey monkeys (SIVsm) on more than one occasion. Reproduced from Rambaut *et al.*, 2001.

The term 'sub-subtype' is used to classify distinctive lineages that are closely related to a particular subtype, but between which there is insufficient genetic difference to justify partitioning into a further subtype division. These include A1 and A2, and F1 and F2 within the subtype A and F groupings (Robertson *et al.*, 2000). Whilst HIV subtypes do associate with certain geographical locations, given the ease of intercontinental travel, some locations such as large cities are the foci for multiple circulating subtypes of HIV-1. The presence of the overall high number of HIV-1 subtypes co-circulating in the Democratic Republic of Congo is suggestive that this region is the epicentre of HIV-1 group M (Vidal *et al.*, 2000b).

HIV-2 is distinct from HIV-1 as, whilst HIV-1 is closely related to SIVs from chimpanzees, HIV-2 is closely related to SIV isolated from sooty mangabeys (Figure 1.6). Like HIV-1, however, HIV-2 has also been classified into sub-groups based on genetic variation, in this case referred to as groups A-G. No sooty mangabey virus with a sequence falling within the HIV-2 groups A, B, C, F or G has been found, but within groups D and E, sooty mangabey viruses have been sequenced which are very similar to HIV-2 virus sequences. It thus appears that each group of HIV-2 represents at least one separate sooty mangabey to human transmission event (Gao *et al.*, 1992).

Following the description of HIV-1 subtypes it was recognised that intersubtype recombinants could develop, for example subtype E (*env* gene) recombined with subtype A (*gag* gene). Such viruses with a recombinant lineage, which are fully transmissible, pathogenic and appear to play an important role in the HIV pandemic, are formally classified as circulating recombinant forms (CRFs) (Cornelissen *et al.*, 2000). All members of a specific CRF group share an identical mosaic structure as they appear to be descended from the same initial recombination event(s). To date 16 CRFs have been formally recognised, the specific recombination characteristics for each CRF shown in the published reference sequences (Table 1.3).

Table 1.3 Summary of circulating recombinant forms.

NAME	REFERENCE STRAIN	SUBTYPES	REFERENCE
CRF01_AE	CM240	A, E	Carr *et al.*, 1996
CRF02_AG	IbNG	A, G	Howard and Rasheed 1996
CRF03_AB	Kal153	A, B	Liitsola *et al.*, 1998; Lukashov *et al.*, 1999
CRF04_cpx	94CY032	A, G, H, K, U	Nasioulas *et al.*, 1999; Paraskevis *et al.*, 2001
CRF05_DF	VI1310	D, F	Laukkanen *et al.*, 2000
CRF06_cpx	BFP90	A, G, J, K	Oelrichs *et al.*, 1998; Montavon *et al.*, 2002a; Vidal *et al.*, 2000a
CRF07_BC	CN54	B', C	Rodenburg *et al.*, 2001
CRF08_BC	GX-6F	B', C	Piyasirisilp *et al.*, 2000, Rodenburg *et al.*, 2001.
CRF09_cpx	96GH2911	not yet published	McCutchan *et al.*, 2000
CRF10_CD	TZBF061	C, D	Koulinska *et al.*, 2001
CRF11_cpx	GR17	A, CRF01_AE, G, J	Paraskevis *et al.*, 2000, Montavon *et al.*, 2002b
CRF12_BF	ARMA159	B, F	Carr *et al.*, 2001, Thomson *et al.*, 2002
CRF13_cpx	96CM-1849	A, CRF01_AE, G, J, U	Wilbe *et al.*, 2002
CRF14_BG	X397	B, G	Delgado *et al.*, 2002
CRF15_01B	99TH.MU2079	CRF01_AE, B	Tovanabutra *et al.*, 2003
CRF16_A2D	KISII5009	A2, D	Gao *et al.*, 2001

Source: http://www.hiv.lanl.gov/content/hiv-db/HelpDocs/subtypes-more.html#HIV-2

1.5.2 HIV evolution

It is generally accepted that the diversity within a viral population is a reflection of that virus' natural history. In the case of HIV, with an evolution rate of nearly 2.4 x 10^{-3} substitutions per base pair per year (Korber *et al.*, 2000), this virus has the capacity to undergo genomic diversification at a rate 5–6 orders of magnitude faster than mammalian genes (Kumar and Subramanian, 2002). This feature makes HIV one of the fastest evolving organisms known to date, and makes the process of understanding its natural history through the anaylsis of gene sequences a unique challenge. The combination of extreme viral genetic plasticity with a complex combination of conflicting evolutionary forces is expressed via viral molecular adaptation and random genetic drift. The natural history of HIV, as told in its genes, is hence inextricably tied to features of both host and virus biology.

1.5.2.1 Mechanisms of generating viral variation

The relative contributions of selection versus stochastic processes in driving viral diversification largely remain unresolved. The mechanistic processes that result in sequence variation, regardless of the driving force, however, are well established.

Mutation and replication rate

As described in section 1.3, the replication of HIV-1 involves the reverse transcription of the viral genome into a DNA copy that integrates into the host cell genome. The viral reverse transcriptase, however, lacks proofreading capabilities due to a lack of 3'-5' exonuclease activity. For HIV-1, therefore, misincorporation, deletion, insertion or duplication of nucleotides occurs during the process of reverse transcription at a frequency of 10^{-4} to 10^{-5}. Together with an *in vivo* virus production rate of over 10^{9} virions per day, the persistence of viral infection and the vast number of people infected, this error frequency gives HIV-1 the potential for enormous expansion of viral diversity. Within the viral genome it is also the case that there is a profound bias toward G to A mutations, known as hypermutation (Vartanian et al., 1991). The reason for this has been elucidated in the last few years with the characterisation of the host anti-viral protein APOBEC3G (Sheehy *et al.*, 2002), which is a member of a protein family that edits DNA/RNA by deaminating cytosine to yield uracil (Harris *et al.*, 2003, Lecossier *et al.*, 2003, Mangeat *et al.*, 2003, Zhang *et al.*, 2003). In the absence of the HIV accessory protein Vif, ABOBEC3G is packaged into virus particles and carried forward

into the next infection, where minus strand reverse transcribed DNA is affected by its action resulting in complete disruption of coding capacity and degradation. The viral protein Vif counteracts the action of APOBEC3G by binding it, inducing its ubiquitination and subsequent proteosomal degradation – precluding its incorporation into new viral particles and hence its anti-viral affects (Mariani *et al.*, 2003, Marin *et al.*, 2003, Sheehy *et al.*, 2003, Stopak *et al.*, 2003, Yu *et al.*, 2003). The existence of such anti-viral proteins, now known to be a powerful mechanism in vertebrate innate immunity (Harris and Liddament, 2004), has thus been imprinted on the HIV genome.

Recombination

An additional mechanism by which HIV sequence diversity may be generated is recombination. This process requires the co- or super-infection of a single host cell by more than one viral particle, during which one copy of each parental genome is encapsidated into a heterozygous virion. When infection by this virus occurs, a recombinant genome is generated as a result of RT switching between alternative genomic templates (Goodrich and Duesberg. 1990). The viral RT is highly recombination prone, with an estimated 3 recombination events occurring per genome per round of replication (Jetzt *et al.*, 2000). As a result, at least 10% of the HIV-1 strains circulating to date result from recombination among distinct viral subtypes (McCutchan, 2000), and recombination has now been detected at all phylogenetic levels: among primate lentiviruses, among HIV-1 groups, among subtypes and within subtypes (Rambaut *et al.*, 2001).

1.5.2.2 Driving forces behind HIV sequence variation

As outlined in section 1.5.1, sequence variation within HIV has been quantified to a certain extent by the segregation of available sequence data into groups, subtypes and sub-subtypes. When considering this classification in the context of the driving forces behind sequence variation, however, one must remember that it is entirely artificial - based on sequence divergence alone. In asking why the subtypes have arisen, therefore, because this classification is not based upon phenotypic divergence the main explanation given is that it is likely that the different HIV-1 subtypes as we know them have arisen due to 'founder effects' in distinct populations (Rambaut *et al.*, 2001). In the case of subtype B in gay men in the US for example, introduction of this genetic variant some time in the late 1970s/early 1980s was associated with high rates of

partner exchange and risky behaviour: the ideal epidemiological conditions for transmission. This means that subtype B rapidly expanded and as a consequence has long been associated with this risk group and geographical location (amongst others). As there is no evidence to suggest that subtype B is more readily transmissible by unprotected anal sex than any other viral subtype, the probability that founder effect has resulted in the prevalence of HIV subtype B in this situation seems high. It is generally accepted that it is unlikely that fitness differences between subtypes determine subtype structure and distribution.

In reality it is also likely that there are HIV variants in existence that would fill the 'sequence gaps' between the subtypes as we know them, but incomplete sampling for now leaves us with this HIV classification. The fact remains, however, that whilst it is unlikely that biological differences between HIV genetic variants initiated their introduction and consequent spread in different host populations, the historical presence of different viral subtypes in genetically very different host populations may have modelled, and may in the future continue to model, their genetic and thus potentially also their biological properties. The question as to whether host adaptation within different populations results in selection of viruses with different biological characteristics is a contentious one.

When considering the generation of HIV sequence diversity, therefore, one must first consider the variety of selection pressures involved. In the case of HIV these include selection within the host during the course of infection, exerted largely by immune pressure, host anti-viral factors and antiretroviral therapy (resulting in intra-host evolution), and selection between hosts (resulting in inter-host evolution). The latter is exerted by both behavioural and genetic diversity within the host population and is strongly effected by selection during transmission. It is because of these differences between the driving forces for inter- and intra-host evolution that the former is not merely a reflection of the latter, on a population scale.

Intra-host evolution

As mentioned, soon after infection HIV is subjected to strong, non-random selection pressures within the host. In such adverse environmental conditions, adaptation takes place rapidly and Darwinian evolution is played out in HIV populations within infected

individuals (Darwin, 1859). Viruses successively fix mutations that enable evasion of immune responses (*env* gene, Richman *et al.*, 2003) and drug selection (*pol* gene, Pillay *et al.*, 2000). Despite potentially decreased fitness these latter mutations often remain in the absence of therapy, most of them becoming fixed in the viral population due to the adaptive benefit they confer. Furthermore, additional changes occur which can compensate for this loss of replicative fitness in drug-resistant mutants (Quinones-Mateu, 2002). It is widely accepted, however, that despite the strong influence of positive selection in intra-host evolution, allele fixation in HIV populations is often also a stochastic process driven by genetic drift, where survival of lineages on the basis of pure chance rather than fitness does take place (Leigh-Brown and Richman, 1997). In this situation the effective viral population size (virus that can contribute to the next generation) is important, as in large populations the abundance of viral variants is subject to fitness competition, while in a small population the fate of mutants is more sensitive to the influence of random events, independent of their fitness. It is not agreed, however, whether intra-host effective population size of HIV-1 is large or small and thus what the relative influences of positive and stochastic processes of selection are. Considering the 'lifetime' of an average HIV infection, it seems most likely that both are important in governing the evolution of HIV within an individual over time, with positive selection dominating under the heavy barrage of the intra-host environment, whereas the drop in effective population size resulting from seroconversion, drug therapy, or immune reconstitution may increase the role of stochastic forces.

Inter-host evolution

Given that HIV-1 intra-host evolution is shaped by the gain and loss of advantageous and disadvantageous mutations respectively, the inter-host evolution of HIV shows little evidence of being driven by positive selection. It has been shown that host-to-host transmission of HIV-1 is accompanied by a loss of genetic diversity, through a bottleneck effect (reviewed in Rambaut *et al.*, 2001). In this case, only a small subset of the donor's viral population is successfully passed on to the recipient, yielding a very homogenous population during primary infection (Delwart *et al.*, 2002), which may not represent either the majority or the most fit viral sub-population from the donor. This bottleneck, however, does depend on population size and transmission route (Dickover *et al.*, 2001). The impact of stochastic processes on inter-host viral evolution is further

enhanced by host behavioural factors. In addition to the bottleneck of the transmission event itself, the vast differences between rates of partner exchange on an individual basis further enhance the process of genetic drift. That is, viral strains with advantageous mutations may be transmitted successfully to a new host, but that host may be one with a very low rate of partner exchange. Consequently, this move in the direction of positive selection is halted. Conversely, a particularly unfit virus may find itself in the situation of infecting a person with a high probability of activity resulting in transmission. Hence, the processes at work in relation to transmission of HIV by no means guarantee the propagation of the most fit virus, and an overall increase viral fitness at the population level.

During the process of infection, as described, there is a great pressure exerted by the host resulting in positive selection of advantageous mutations and out-growth of viruses best adapted to that host. The main point of controversy when considering this process of positive selection is, within a specific host population of similar genetic composition i.e. ethnic origin, is it likely that the viruses within these populations will experience intra-host positive selection in the 'same direction'. It has been argued that advantageous mutations, such as those conferring escape from the activity of cytotoxic T-lymphocytes (CTL), might not appear until relatively late in infection (Goulder *et al.*, 1997). If these positively selected escape mutants do not appear until after most individuals have transmitted their virus, natural selection of this advantageous mutant will not occur at the population level. Furthermore, it has been shown that in some cases when transmission of CTL escape mutants does occur, if the virus is transmitted to an individual with a non-matched human-leukocyte antigen (HLA) background, then the CTL escape mutations are in fact deleterious and are removed by purifying selection (Leslie *et al.*, 2004).

There is, however, some evidence that the influence of host genetics in different populations has indeed moulded the genomes of the dominant viral subtype in circulation, and that to a certain extent positive selection of HIV does occur inter-host, at the population level and can influence the biological properties of different HIV subtypes. Specifically, a recent study investigated the likely selective pressures that have affected HIV-1 group M *env* sequences and their diversification from the original group M founder virus (Travers *et al.*, 2005). Each individual subtype was compared

to all other group M subtypes in an attempt to identify amino acid sites whose evolutionary history appeared to be unique in terms of selective constraints for that subtype. The authors showed that there was considerable heterogeneity of evolutionary mechanisms among the different subtypes and identified groups of sites that were subject to different selective constraints in the lineages leading to each subtype. One category of amino acids were identified that had undergone positive selection for subtypes C, F1 and G, but had undergone purifying (negative) selection in all other subtypes. Amino acid sites within subtypes A and K formed the second category identified, which had undergone purifying selection in these lineages, but had undergone positive selection in all others. Whilst the functional consequences of these differences have not been shown, it was hypothesised that the heterogeneous selection pressures exerted by different host populations within which the HIV-1 subtype lineages diverged, may account for the various levels of fitness that have been observed in the group M subtypes. For example, amino acid residues involved in CD4 binding, sites involved in CCR5 to CXCR4 coreceptor switch, sites within the flexible linker of gp41 (important for tethering gp41 to the transmembrane segment, coreceptor binding and host cell entry) and sites corresponding to known neutralising antibody epitopes, had all undergone differential selection processes. Whilst a small study, this work is novel in that it considers for the first time that the evolution of the HIV group M subtypes, under different selection pressures in different host populations, may have resulted in subtype-specific genetic differences and thus potentially subtype-specific biological differences.

1.5.3 The Importance of Genetic variation: of the Host and the Virus

Host

The importance of host genetic variation in determining the outcome of HIV infection has been studied for some time and, whilst well established, the full extent of this has yet to come to light. The majority of information relates to human genetic factors that may affect the probability of acquisition of infection, rather than those that may affect the probability of propagation of infection. It is difficult, however, to distinguish the affects of the infecting donor and infected recipient in terms of transmission, although in some cases the relative contributions are clear. Table 1.4 shows a number of associations between human gene polymorphisms and susceptibility to HIV-1 infection, and subsequent disease progession, which have been characterised.

Table 1.4 Associations between human gene polymorphisms and susceptibility to HIV-1 infection.

Gene, Marker or Variant	Transmission mode	Ethnicity of subjects (location)	Subtype	Effect on risk
CCR5				
Δ32 (HHG*2) *[2]	MSM, IDU, Het, Tx,	White (multiple)	B	Decrease
Homo/heterozygous	MTC*[1]	White (multiple)	B	None
Promoter HHE (P1/P1) *[3]	MSM, IDU, Het	White, African American (US)	B	Increase
Homozygous	MTC	White (Argentina)	B	Increase
HLA				
Concordance at HLA I loci or	MTC	African (Kenya)	A	Increase
HLAB	Het	White (UK)	B	Increase
	MTC	African American (US)	B	Increase
	Het	Zambian	C	Increase
	MTC	African (Kenya)	A	Decrease
A2/6802 and A0206/6802*[4]	MSM	White (US)	B	Decrease
CHEMOKINE GENES				
CCL3LI – segmental duplication of gene (more copies relative to same members of ethnic group)	Multiple	Multiple	Multiple	Decrease

*[1] Abbreviations for transmission mode: MSM, men who have sex with men; IDU, intravenous drug user; Het, heterosexual; Tx, transfusion or blood product; MTC, mother to child.
*[2] The Δ32 (HHG*2) variant is a 32 bp deletion in the region coding for a transmembrane domain of the CCR5 coreceptor. Homozygosity affords almost complete protection, and heterozygosity slight protection.
*[3] This CCR5 haplotype enhances risk of infection (for CCR5 haplotype nomenclature see Gonzales *et al.*, 1999).
*[4] Designations are for HLA supertype (see MacDonald *et al.*, 2001)
Modified from Kaslow, Dorak and Tang, 2005.

Chemokine receptors and their ligands play an important role in modulating susceptibility to HIV-1 infection. The importance of the CCR5 receptor is outlined in Table 1.4, and in addition to this other receptors and their ligands have been implicated. For example, coding polymorphisms in the CCR2 gene have been associated with a favourable prognosis during infection, as have changes in the CCL5 gene that encodes RANTES (the ligand for CCR5). Changes in the latter in different populations, however, have also been shown to confer some disadvantage to HIV-exposed individuals. CXCL12, which encodes SDF-1, (the ligand for CXCR4) has also been implicated in modulating host susceptibility to infection, but data are less clear in this case. As shown in Table 1.4, it is not only chemokine gene polymorphism that can modulate susceptibility to infection. The recent description of the effects of segmental duplication of the CCL3LI gene, which also encodes a CCR5 ligand, has further

expanded our understanding of the host's involvement in directing the process of HIV infection (Gonzalez *et al.*, 2005).

The central importance of the MHC class I antigen processing and presentation system, in terms of HIV-1 pathogenesis, however, is not in doubt. Because HLA class I alleles differentially bind viral peptides, restrict generation of CD8+ CTLs and thus govern response to viral antigens, variation in these alleles within humans greatly varies the ability of different individuals to either promote or impede the initial establishment of infection and indeed its progression. The specific details of MHC variation and its significance in disease are thoroughly reviewed in Kaslow, Dorak and Tang, 2005. It would be inappropriate, however, to mention MHC class I variation and not refer to recent research which has outlined a dominant role of HLA-B in mediating the potential co-evolution of HIV and HLA. As mentioned, the importance of HLA-B in governing disease outcome is well known. In this study, however, preliminary evidence suggests that because of its importance, the rate of diversification of the HLA-B allele is much more rapid than HLA-A and–C loci, in the face of pressure exerted by the presence of HIV. The HIV genome accordingly bears the evolutionary scars of repeated CTL escape from HLA-B induced immune pressure. Whilst not proving definitively that a process of HLA and HIV co-evolution is at work, in the light of this study a picture is emerging which is highly indicative of such a process, especially in populations where HIV morbidity and mortality rates are high (Kiepiela *et al.*, 2004).

Virus

Unlike genetic variation in the host population, the importance of HIV genetic variation in relation to disease is much less well understood. The literature on this subject will, however, be thoroughly reviewed in the introduction to Chapter 3 of this thesis. Rather, at this juncture it should be noted that our ability to currently classify HIV-1 sequence variation into subtypes facilitates epidemiological studies of HIV-1. Whilst the significance of viral genetic variation (manifest as subtypes) if any, is not clear, their existence enables detailed studies of HIV transmission dynamics, on a global scale, to be performed. In recent years, subtyping of HIV sequence data produced for clinical resistance testing, for example, has indicated an increase in subtype diversity in the UK (Parry *et al.*, 2001, Tatt *et al.*, 2004). This suggests that the 'import' of new genetic variants into the UK population from distinct geographical locations is taking place, and

gives both researchers and clinicians an important perspective on the transmission dynamics within this population: where incident infections are arising and which sections of the population are most at risk, for example (Tatt *et al.*, 2004). Furthermore, whilst classification of HIV-1 variation into subtypes in no way describes the totality of HIV-1 sequence variation within populations, it does provide a system that allows the effects of HIV-1 sequence variation to be studied in sufficient numbers to be meaningful.

1.6 Project introduction

The aim of this thesis is to characterise the genotypic and phenotypic features of the different types and subtypes of HIV. The question is asked: are the differences between genotypic variants of HIV manifested phenotypically and if so, do they have biological relevance?

It is true that in considering HIV subtype as a 'risk' factor, many epidemiological studies are flawed in that even if subtype is a true causal factor in disease phenotype, it may be indistinguishable from other potentially multiple factors that contribute to a particular outcome such as faster or slower disease progression. Comprehending these data can be difficult at the best of times, so for this study a conceptually more simplistic approach will be taken. Firstly, a computational method for accurately classifying large amounts of HIV sequence data, by subtype, will be created. In order for large-scale studies to be performed in which viral subtype is related to clinical outcome, this method must be both rapid and accurate and make use of the must abundant source of sequence data, preferably linked to clinical information. Secondly, the phenotypic characteristics of several different HIV-1 subtype isolates and HIV-2 will be established, *in vitro*. The aim of this is to determine if, in a homogenous environment, some viral variants are reproducibly more or less 'fit' than others. Finally, if different viral subtypes grow differently in cell culture the hypothesis may be put that genetically variable strains of HIV will induce different transcriptional profiles within infected cells. If this is true, and these agree in any way with previously published observations regarding differences between HIV subtypes, this could lend weight to the argument that subtype is an important causal factor in the phenotype of HIV disease.

Furthermore, by studying the transcriptome of cells infected with divergent strains of HIV using DNA microarrays, it should be possible to define a core response to the presence of this virus, regardless of variant type. These may indicate the pathways with which HIV must interact in order to complete its life cycle and thus provide novel targets for antiviral intervention.

Chapter 2.0 Development of a HIV subtyping tool

2.1 Introduction

Prior to the mid-1980s and introduction of polymerase chain reaction (PCR) to the laboratory, viruses were mainly identified by growth in cell culture and observation for cytopathic effects (CPE). This method was routinely used as the first means of detection and presumptive identification of the virus in question. Before the introduction of molecular methods, typing and subtyping of viruses was mainly achieved by serotyping: a method that employs antibodies that define antigenic differences. Serologic methods have been used to define major viral groups, or types, such as influenza virus types A to C, parainfluenza virus types 1 to 4, poliovirus types 1 to 3, and herpes simplex virus types 1 and 2 (HSV-1 and -2, respectively) (Arens, 1999). Some virus species, however, such as human cytomegalovirus (HCMV) and measles virus, cannot be divided into different types or subtypes based on serology because significant antigenic differences do not exist (Arens, 1999). Thus, whilst serotyping has been useful for making relatively large distinctions between viruses it does not have the ability to distinguish certain viral types, or between individual isolates within a serotype. Such fine classification has become possible with the development of molecular methods of virus classification.

PCR and the ability to amplify single copies of DNA or RNA molecules to detectable levels, together with DNA sequencing, has revolutionised the process of detection and characterisation of viral genomes. As a result, laboratories whose interest is the diagnosis and epidemiology of viral diseases have largely converted from conventional techniques of virus detection and identification, to more rapid, novel, and sensitive molecular methods. It is possible, however, that in doing so useful phenotypic characteristics of viruses are not being determined and this may impair studies of virus biology. This concern could be applied to many viruses, but is perhaps an inevitable outcome of the paradigm shift of virus classification: from a discipline solely dependent on the observation of viral phenotypic traits (*in vivo* or *in vitro*), to one in which phenotypic information has become secondary to the detail contained in the viral genome sequence.

A summary of molecular methods for characterisation of viruses is presented in Table 2.0, and a selection of the viruses to which these methods have been applied is summarised in Table 2.1. Multiple methods exist to facilitate the subtyping of viruses, ranging in their complexity, efficiency, spectrum of detection and suitability for different viruses. Of all the methods, however, it is DNA sequencing and methods that employ sequence-based hybridisation, especially microarray and chip-based techniques, which currently dominate the field of viral subtyping or more specifically, genotyping.

Table 2.0 Molecular methods for characterisation of viruses.

Method	Restriction enzyme	Probe	Basis for distinction	Level of resolution	Advantage(s)	Disadvantage(s)
Nucleotide sequencing	No	No	Nucleotide sequence	Single genome (if cloned)	Wide applicability; can identify single nucleotide mutations	Technically complex; produces large amounts of data; automated sequencing requires expensive equipment
Solid phase plate assay (SPPA)	No	Yes	Nucleotide sequence	Subtypes	Simple; commercially available	Developed for HIV. Does not detect circulating recombinant forms
Microarray and BioChip	No	Yes	Nucleotide sequence	Single genome (if cloned)	Wide applicability; can identify single nucleotide mutations	Technically complex; produces large amounts of data; requires expensive equipment
RFLP analysis	Yes	No	Restriction sites	Subtypes	Simple; wide applicability	Samples only the restriction sites; difficult to use with RNA viruses
Southern blotting	Yes	Yes	Restriction sites, probe sites	Subtypes	Wide applicability; ability to analyse complex genomes	Samples only the restriction sites and probe sites; technically complex; may require radioisotopes
Oligonucleotide fingerprint analysis	No	No	RNase T_1 cleavage sites	Subtypes, quasi-species	Directly applicable to RNA viruses; can detect point mutations	Involves complex electrophoresis procedure
RH	No	Yes	Probe sites	Subtypes	Simple; commercially available	May not correctly identify all subtypes of HCV
DEIA	No	Yes	Probe sites	Subtypes	Simple; commercially available	May not correctly identify all subtypes of HCV
RNase protection assay	No	Yes	Probe sites	Subtypes	Readily applicable to RNA viruses	Requires a radioactive probe; technically difficult
SSCP analysis	No	No	Mobility differences	Subtypes	Detects a few mutations in a large number of bases; simple procedure	Identifies the presence but not the location of mutations
HMA	No	No	Mobility of hetero- vs homo-duplexes	Quasi-species	Can visualize a large number of quasispecies	Cannot distinguish quasispecies with <2-3% nucleotide differences; technically difficult and complex
Genome segment length polymorphism analysis	No	No	Segment length	Subtypes	Easy technique for use with segmented genomes	Cannot detect mutations (detects only variations in segment length); applicable only to viruses with segmented genomes

Modified from Arens, 1999.

Table 2.1 Application of molecular subtyping methods to specific viruses.

Virus	Method(s)	Reference(s)
Adenovirus	RFLP analysis	Hierholzer, 1992, Hierholzer *et al.*,1988, Li *et al.*, 1996
CMV	RFLP analysis, PCR-RFLP analysis, Southern blotting, glycoprotein B genotyping	Huang *et al.*, 1980a,b, Kilpatrick *at al.*, 1976, Spector *et al.*, 1982, Tolpin *et al.*, 1985, Winston *et al.*, 1985, Chern *et al.*,1998
Dengue virus	Nucleotide sequencing	Deubel *et a.*, 1993
Epstein-Barr virus	RFLP analysis, analysis of termini	Falk *et al.*, 1995, Katz *et al.*, 1988, Rabb-Traub and Flynn, 1986, Sidagis *et al.*, 1997
Enterovirus	RFLP analysis, RT-PCR-RFLP analysis, oligonucleotide fingerprint analysis, SSCP analysis, nucleotide sequencing	Fujioka *et al.*, 1995, Kuan, 1997, Minor *et al.*, 1982, Mulders *et al.*, 1995, Nottay *et al.*, 1981
HBV	Southern blotting, SSCP analysis, PCR-SSCP analysis, nucleotide sequencing, PCR/RFLP	Hardie *et al.*, 1996, Hasegawa *et al.*, 2004
HCV	SSCP analysis, PCR-SSCP analysis, RH, DEIA, nucleotide sequencing	Biasin *et al.*, 1997, Castelain *et al.*, 1997, Lareu *et al.*, 1997, Lee *et al.*, 1997, Le Pogam *et al.*, 1998
HIV	HMA, HTA, nucleotide sequencing, SSPA	Delwart *et al.*, 1993, McNearney *et al.*, 1992, Nelson *et al.*, 1997, Tatt *et al.*, 2000, Ou *et al.*, 1992, Gonzalez-Villasenor *et al.*, 2000
HSV	RFLP analysis, PCR-RFLP analysis, Southern blotting	Buchman *et al.*, 1978, 1979
Influenza virus	RNase protection assay, HMA, nucleotide sequencing, DNA chip, microarrays	Lopez-Gallindez *et al.*, 1988, Subbarao *et al.*, 1998, Zou, 1997, Kessler *et al.*, 2004, Li *et al.*, 2001
JC virus	Nucleotide sequencing	Agostini *et al.*, 1997
Measles virus	RT-PCR-RFLP analysis	Katayama *et al.*, 1997
Parainfluenza virus	Serotyping, sequencing	Beraud *et al.*, 1984, Komada *et al.*, 1991
Parvovirus B19	SSCP analysis, PCR-SSCP analysis	Kerr *et al.*,1995
RSV	RNase protection assay, PCR-RFLP analysis, sequencing	Cristina *et al.*,1991, Storch *et al.*, 1991
Rhinovirus	Serotyping, nucleotide sequencing	Cooney *et al.*, 1982, Duechler *et al.*, 1987, Hughes *et al.*, 1988, Rivera *et al.*, 1988
Rotavirus	Genome segment length polymorphism analysis	Arens and Swierkosz, 1989, Dolan *et al.*,1985, Gaggero *et al.*, 1992
Small round-structured viruses (Norwalk-like viruses)	Southern blotting, nucleotide sequencing	Ando *et al.*, 1995, Levett *et al.*, 1996, Yamazaki *et al.*, *1996*
Varicella-zoster virus	RFLP analysis, long PCR-RFLP analysis	Adams *et al.*, 1989, Takayama *et al.*, 1996

Modified from Arens, 1999.

Today, after approximately 30 years of sequencing and databasing of sequence data, the NCBI database Entrez Genomes contains 1864 reference sequences for 1344 viral genomes and 36 reference sequences for viroids. Not only are these data central to our ability to investigate and understand the molecular biology of viruses, but also provide an invaluable resource in the study of the molecular phylogeny of viruses and their evolution. The availability of well-characterised reference sequences for many viruses, together with thousands of sequences of viral isolates deposited daily around the world,

has enabled the construction of virus phylogenies using phylogenetic techniques, and the categorisation of virus families, type and subtypes, based on sequence data.

2.1 1 Phylogenetics and subtype classification

After sequencing, the cornerstone to any genotypic classification - be it of an entire viral genome or a short fragment of a particular gene, are the phylogenetic methods used to determine the relationships between the sequences being studied.

Phylogenetics is a mathematical analysis of sequence similarity that determines the evolutionary relationships between sequences based on derived similarity or difference. The task of molecular phylogenetics is to convert information in sequences into a phylogram, or phylogenetic tree. Three of the many phylogenetic analysis methods currently in widespread use were used in the course of this work; hierarchical clustering of sequence identity matrices, neighbour joining (NJ - implemented from within CLUSTALW) and maximum parsimony (MP). The three methods fall into two phylogenetic categories, distance methods (hierarchical clustering, NJ) and discrete methods (MP). Common to all, however, is the fact that the basis of determining the relationship between sequences relies on determining their similarity. To determine overall maximal similarity between two or more sequences, they must first be correctly aligned.

2.1.1.1 Sequence alignment

The most fundamental aspect of sequence analysis is pair-wise sequence alignment, during which two sequences are compared to determine if there is a relationship between them that is unlikely to have occurred by chance. This therefore requires the sequences to be accurately aligned, a way of scoring the similarity between the aligned sequences, and a method to determine the significance of the alignment relative to a measure of a chance alignment of the same score (Holmes and Durbin, 1998). Most common scoring methods are additive, where each position within an alignment is assumed to be independent of other positions. A score is derived from each position in the alignment, which is then summed along the length of the sequence. Matches score more highly than mismatches, producing an overall higher score where the sequences aligned are more similar. In addition, one or both sequences may contain regions that are absent from the other. Such insertions or deletions do not score. In many sequence

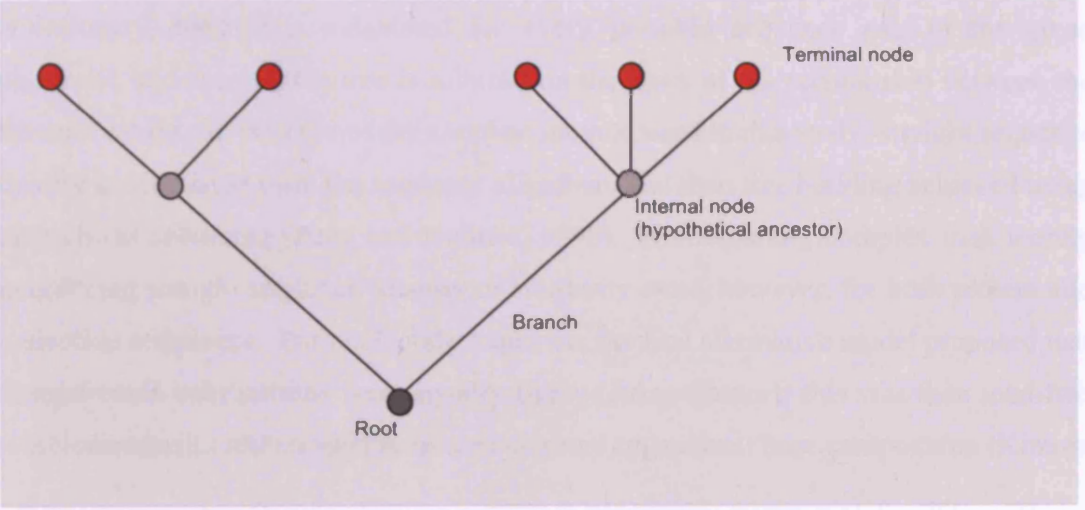
alignment tools, however, the most simplistic match/mismatch score of 1/0 (respectively) is replaced by a scoring system that attempts to quantify the degree of mismatch between two positions in the sequence, giving greater insight into the similarity, or difference, between aligned sequences (described in section 2.1.1.4).

Pair-wise alignment is often used as the basis for a more useful form of sequence analysis, the multiple alignment. A common implementation of the multiple sequence alignment is CLUSTALW (Thompson, 1994). CLUSTALW generates a distance matrix of alignment scores from all sequence pairs, then uses a clustering algorithm (NJ - Saitou and Nei, 1987) based on those distances to construct a guide tree to relate those sequences. Nodes within the tree are then aligned in order of decreasing similarity, starting with the two most similar sequences. The result of this process is a set of sequences aligned to each other.

2.1.1.2 Phylogenetic trees

A phylogenetic tree is a mathematical structure that is used to model the proposed evolutionary history of a group of sequences or organisms. The pattern of historical relationships is the phylogeny or phylogenetic tree, which can be estimated using a variety of methods. Regardless of the methods by which the tree itself is generated, however, it always consists of nodes, connected by branches. Terminal nodes represent sequences for which data are held, whereas internal nodes represent hypothetical ancestors that can be inferred by examination of the data in hand. The ancestor of all the sequences in the tree is the root (Figure 2.0).

Figure 2.0 A simple rooted tree and associated terms (from Page and Holmes, 2001)



Various kinds of information may be associated with the nodes and branches of a tree, depending on the method used to construct it. Most trees attempt to estimate the amount of evolution that takes place between each node on the tree, translated in the tree topology as branch length.

Phylogenetic trees may be either rooted, or unrooted. A rooted tree has a node identified as the root from which all other nodes descend, and hence has direction. The closer a node is to the root of the tree, the older it is in evolutionary time. Rooting a tree, therefore, gives a clear picture of the ancestor-descendant relationships within the sequences studied. Typically, a sequence that represents an ancestor, or distant relative, of the sequences being studied is chosen as the root. If studying HIV-1 group M sequences, for example, a HIV-1 group O sequence could be used as the root. Un-rooted trees lack a root and as such do not infer the same relationships between sequences as when displayed in a rooted tree. In un-rooted trees the proximity of nodes can only be interpreted as indicating the similarity or distance of the sequences at those nodes, to each other. There is therefore no directionality to un-rooted trees.

2.1.1.3 Distance methods

Distance based phylogenetic methods rely on matrices of pairwise sequence distances, the most simple sequence distance being the fraction of identical residues (nucleotide or amino acid) between two aligned sequences. Given an alignment of sequences, distance methods first convert these aligned sequences into a matrix of pair-wise genetic distances. A measure of dissimilarity between sequences is calculated for every pair of sequences in the alignment as the fraction of positions in which the two sequences differ. The distance method process therefore takes place in two stages; first, the evolutionary distance is calculated for every possible sequence pair in the given alignment, and second, the tree is inferred on the basis of the relationship between the distance values. In the case of the simplest method used in this study, straight sequence identity is calculated from the sequence alignment and then tree building achieved using hierarchical clustering (Page and Holmes, 1998). Models more complex than merely considering straight sequence identity or similarity exist, however, for both protein and nucleotide sequences. For nucleotide sequence, the first alternative model proposed that all nucleotide substitutions were equally likely (Jukes-Cantor); this was then modified to include transition/transversion frequencies and organismal base composition (Kimura

et al., 1980). The General Reversible model (Yang *et al.*, 1994) allows for unequal frequency between each of the 6 pairs of nucleotide substitutions. In this study, however, we did not attempt to use a more complex model than simply measuring sequence identity – rather two separate tree building methods were employed to represent the same distance matrix (hierarchical clustering and NJ).

Distance methods of determining molecular phylogeny are the most simple as they assume that the genetic distance between sequences being compared can be applied to a tree to directly reconstruct the evolutionary history of those sequences. It is rare, however, that these distances match exactly tree metrics. As such, 'goodness of fit' methods seek the tree that best accounts for the 'observed' distances. Alternatively, the tree whose sum of branch lengths is minimal (i.e. constrained to find the tree with the shortest evolutionary pathway) may be sought ('minimum evolution'). In this study, hierarchical clustering and NJ algorithms were used to build trees based on sequence identity. The former is the simplest method, and is not commonly used to represent phylogenetic relationships between sequences. NJ - the second distance based phylogenetic method used (Studier and Keppler, 1988), however, is one of the most popular methods for tree building and is implemented within the CLUSTALW algorithm. It combines computational speed with ease of implementation and employs a minimum evolution strategy on the basis of which a single tree is built by gradually finding neighbours exhibiting the minimum genetic distance. NJ, however, is a heuristic method for estimating the minimum evolution tree and does not guarantee to find the most likely tree. The main criticism of distance methods such as these, therefore, is that by summarising a set of sequences by a pair-wise matrix, information that could be used to better understand the relationship between those sequences, is lost.

2.1.1.4 Discrete methods

Discreet methods for construction of phylogenetic trees attempt to avoid the loss of information inherent in distance-based methods, when sequences are converted to distances. The two main discrete methods are Maximum parsimony and Maximum likelihood.

Maximum parsimony, the third and most complex phylogenetic method used in this study, is a discrete tree building method that takes a multiple alignment of sequences as

input and chooses the tree (or trees) that requires the fewest evolutionary changes, or mutations. That is, starting from an initial topology, the maximum parsimony algorithm infers the minimum number of mutations required to justify all nodes of the tree at every sequence position. The process is repeated for all theoretically possible tree topologies and the tree requiring the minimum number of evolutionary changes, called the minimum tree, is selected as the best tree. It is the case, however, that by constraining the optimal tree to one that invokes the fewest possible evolutionary changes, the result may not necessarily reflect the minimal evolution between two terminal nodes. Given that all evolutionary changes between two sequences may not be visible, such as multiple substitutions at the same nucleotide position, the dissimilarity between two distantly related sequences might be underestimated in a parsimony tree. Despite this, we used an MP method in this study: PHYLIP protpars (Felsenstein, 1989). This program counts amino acid substitutions in terms of the mutations at the nucleotide level that would be required to go from one amino acid to another.

As an alternative, the discrete method maximum likelihood chooses the tree (or trees) that, of all trees, is the most likely to produce the observed data. It is computationally significantly more expensive than any methods described above and, whilst perhaps constituting the best method available to reconstruct a molecular phylogeny, it has not been used in this study.

2.1.2 Viral subtyping

In essence, molecular characterisation and subtype classification of viruses relies upon the basis that viral genes may vary (mutate), within the functional constraints of the proteins they encode. PCR amplification and sequencing allows small sections of the viral genome, including small variable regions that define viral subtypes within more conserved genome regions, to be compared. In addition, with the development of 'long PCR' methods entire viral genomes may be amplified and used to determine subtype specific classification, which in this case may include traits held in un-translated regions, unique as they may include random mutations that persist in the absence of immunological pressure. Furthermore, with whole-genome analysis recombinant genomes, composed of two or more different subtype-derived fragments, may be detected.

In certain cases, however, small regions of a viral genome sequence may be the only genetic information available. For example, clinical isolates of HIV routinely have only protease (PR) and part of reverse transcriptase (RT) amplified and then sequenced (PR-RT). The products of these genes are drug targets, which are sequenced to determine antiretroviral resistance. Such regions were not traditionally considered suited to subtyping due to low levels of sequence variation. However, several studies have suggested that it is possible to distinguish between subtypes based on the sequence variation within *pol* (Pasquier *et al.*, 2001, Njouom *et al.*, 2003, Gale *et al.*, 2004). In this situation, therefore, a readily available resource may be exploited for HIV subtyping.

2.1.2.1 HIV subtyping

The molecular classification of HIV was previously described (Chapter 1.0). In short, maximum likelihood phylogenetic analysis of *env* gene sequences revealed three lineages; M (major), O (outlier) and N (non-M/O) (Simon *et al.*, 1998). Within group M, 9 major subtypes were identified, termed A-D, F-H, J and K, and within the subtype A and F groupings the sub-subtypes A1 and A2, and F1 and F2 were characterised (Robertson *et al.*, 2000). Several circulating recombinant forms have also been described. HIV subtyping methods predominantly rely on phylogenetic analysis of HIV-1 nucleotide sequences, often of the envelope gene. For more high throughput analysis, however, three web-based genotypic subtyping tools are available, at NCBI (Subtyping Tool), Stanford (HIV-SEQ) and Los Alamos (SUDI) websites.

The NCBI Subtyping tool uses the BLAST algorithm to score local sequence similarity along a sliding window of nucleotide sequence to match the most similar subtype to a query sequence (Rozanov, 2004), and any part of the HIV genome may be analysed. Alternatively, the Stanford tool calculates genetic distance between the query sequence and a set of reference sequences, but will classify only PR-RT sequence. Both of these web based subtyping tools use a limited set of subtype reference sequences that represent only a small fraction of the true sequence variability within each HIV subtype. Both methods also compare the query sequence against each reference sequence independently, and as a consequence they lack the ability to assign subtype based on group properties, or subsets of properties. The use of a BLAST based tool to analyse very highly related sequences (NCBI) is also problematic as the sequence modelling

and statistics associated with BLAST were designed to evaluate the probability of a match compared with random sequence (http://www.ncbi.nlm.nih.gov/BLAST/). This form of random model is therefore inappropriate when comparing clinical HIV-1 sequences with HIV-1 reference sequences, as whilst the lowest BLAST score may indicate the most probable subtype, the scores from other subtypes are equally likely to be significantly similar. Essentially the use of BLAST, in this context, renders the computed statistics invalid. The Stanford subtyping tool, however, provides a better alternative and it is this method which is used to subtype the majority of sequence data produced during resistance testing as the same website provides a facility to simultaneously detect resistance and resistance-associated mutations (http://hivdb.stanford.edu/). Neither the NCBI nor Stanford methods attempt to assign any confidence to the subtype predictions made, however, making assessment of the reliability, sensitivity and specificity of their subtype predictions very difficult. Both methods require nucleic acid as input.

It is clear, therefore, that the subtyping methods available vary in the level of computational and statistical robustness for assigning subtype, the two methods mentioned providing only a basic means of assigning subtype. An alternative method, SUDI, is designed to help determine if a newly defined set of related sequences should most appropriately be considered a new subtype, a new sub-subtype, or part of a previously defined subtype. Ideally, it should be used in conjunction with other tools that will identify recombination and phylogenetic relationships, and as such requires a certain level or pre-processing of HIV sequence (requiring phylogenetic analysis packages and some considerable computing power).

To date, most studies involving the determination of HIV subtypes have used the *gag* and *env* genes for subtyping as their genetic variability is greater than that seen in *pol*. As mentioned previously, however, the more conserved *pol* gene is sequenced routinely in many laboratories to identify antiretroviral therapy (ART) resistance-associated mutations and as such has made nucleotide and amino acid sequences from these regions widely available as material for subtyping HIV-1. As a consequence of recent guidelines (Hirsch *et al.*, 2000; Miller *et al.*, 2000), the practice of sequencing this region of the HIV genome in patients exposed to ART will become increasingly commonplace. The volume of Pol sequences available, or more specifically PR amino

acids 1 – 99 and RT amino acids 1 – 340 (PR-RT), should thus prove valuable in molecular epidemiological studies. It is for these reasons that the aim of the work described in the first section of this thesis was to develop a HIV subtyping tool that could use sequences, produced for resistance testing, as its input.

2.1.2.2 Relevance of HIV genotyping

Epidemiology

The accurate and rapid determination of HIV-1 subtypes is important from an epidemiological standpoint. By monitoring the distribution of subtypes in a population much information can be gleaned about the transmission dynamics of the epidemic in question, its age and the rate of introduction of new sequences. Since the onset of the HIV pandemic in the early 1980s there has been a renewed interest in the emergence and evolution of infectious disease. In the case of HIV, the use of molecular phylogenetics in the reconstruction of the HIV-1, -2 and SIV phylogeny was described in the introduction to this thesis. The use of phylogenetic trees has been central to this process, but their use is not limited merely to understanding the relationships between sequence data in a historical sense. On the contrary, molecular phylogenetic methods – be they reconstruction of a full maximum likelihood tree, or the classification of viral subtypes and analysis of their distribution in a host population – are the cornerstone of contemporary studies of infectious disease, and the development of predictive algorithms of how new and emerging diseases may affect populations in the future.

One of the earliest well documented uses of phylogenetic analyses of HIV sequence data, prior to the inception of the subtype classification, was the study of the 'Florida Dentist'. In 1990 the Centres for Disease Control (CDC) received reports of AIDS in a young woman in Florida whose only risk of HIV infection was that she had previously been treated by a dentist suffering from AIDS. The investigation that followed uncovered a number of the dentist's former patients who were also HIV positive, the inference being that the dentist had also infected these people. Env sequences were amplified from the dentist, the individuals involved and from several individuals living in the local community (not linked to the dentist), and a phylogenetic tree reconstructed (Figure 2.1). This revealed that 5 of a cohort of 7 patients, with no other risk factors for HIV infection, had sequences closely related to those of the dentist. This strongly suggested that the dentist infected them as a result of dental treatment (Ou *et al.*, 1992).

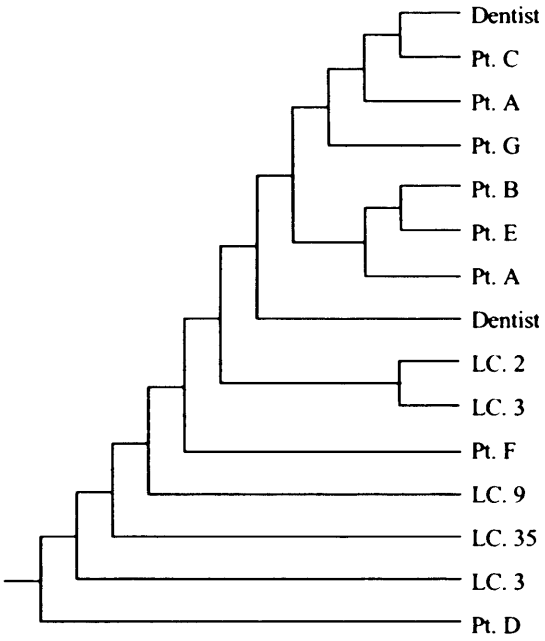


Figure 2.1 Phylogenetic tree relating HIV-1 envelope coding sequences to a HIV-1 infected Florida dentist.
Pt. A-G indicates HIV-1 sequences taken from seven patients (Pt) of the infected dentist. LC (local control) samples were taken from HIV-1 infected individuals living within a 90-mile radius of the dental practice. Two samples of were used from both the dentist and patient A. Reproduced from Ou *et al.*, 1992.

This example shows how sequence information in the form of mutations that had accumulated within patients, between transmission events, could recount the very recent history of its spread. When performed at population or global levels this sort of detailed study of individuals, although complex, allows patterns of transmission and spread of HIV to be monitored: such an approach has been used to model transmission dynamics within the MSM population in the UK (Hue *et al.*, 2005). Such a study however, whilst significant, is immensely computationally expensive and requires highly sophisticated analysis and interpretation. At the population level therefore, whilst less sophisticated, the classification of distinct genetic variants of HIV (subtypes) that are co-circulating and the monitoring of their prevalence and dynamic spread in a population may be valuable for epidemiological studies.

Genotype, phenotype and antiretroviral resistance

With sequence variation, differences in viral infectivity, transmissibility and immunogenicity may exist (Robertson *et al.*, 2000). Whether the various groups, subtypes and recombinant forms of HIV-1 have biological differences, however, is largely unclear and cannot effectively be determined without accurate genotyping of viruses from large numbers of HIV-infected patients and relating this to clinical information.

More clear, however, is the relationship between HIV genetic subtype and natural resistance to antiretroviral drugs (Apetrei *et al.*, 1998, Descamps *et al.*, 1997, Descamps *et al.*, 1998), as well as between subtypes and the efficiency of serological and molecular tests for HIV diagnosis (Apetrei *et al.*, 1996, Simon *et al.*, 1994, Parekh *et al.*, 1999). Major mutations linked to resistance to non-nucleoside RT inhibitors (NNRTIs) can be detected in all HIV-1 group O isolates (A98G and Y181C) and in one group M (subtype J) virus (V108I) (Vergne *et al.*, 2000). In terms of accessory resistance mutations, a study of 142 subtype B or D virus isolates derived from treatment-naïve patients, showed that only 5.6% had no drug resistance associated mutations in the protease gene (Vergne *et al.*, 2000). A phylogenetic study of a group of seroconverters from France that combined *pol* gene sequencing with envelope serology also revealed a diverse range of viruses (Fleury *et al.*, 2003): the protease polymorphisms K20I and M36I were highly prevalent in subtypes CRF02(AG), G, J, F2, and CRF06(cpx). Only three isolates, however, were found to have NRTI- or NNRTI-associated resistance mutations: one subtype D isolate (M184I), one subtype J isolate (T69N, K70R), and one subtype A isolate (K103N). It can therefore be seen that knowledge regarding the subtype of virus with which a patient is infected may have a bearing on treatment choice. The need for such information was clearly illustrated in a report that described, among 30 patients failing nelfinavir-containing therapy as their first protease inhibitor (PI) regimen (19 were infected with HIV-1 subtype B and 11 with subtype G), nelfinavir-associated resistance mutations differed significantly depending on viral subtype (Gomes *et al.*, 2002). The results suggest that the two distinct pathways (D30N and L90M) which lead to nelfinavir resistance may be guided by subtype-specific polymorphisms within the backbone of the protease gene, and that subtype G infected patients could be at much higher risk of developing broad PI cross-resistance when nelfinavir is used as the initial PI. This apparent differential response

Table 2.2 Published studies on drug resistance in persons infected with HIV-1 non-B isolates.

Region	Country	Reference	Main subtypes	Patients
Africa	Cote d'Ivoire	Adje *et al.*, 2001	A, AG, D, G, H	68
	Uganda	Weidle *et al.*, 2002	A, C, D	94
	Gabon	Vergne *et al.*, 2002	A, AG, AE, D, G, H, J	22
	Zimbabwe	Kantor *et al.*, 2002	C	21
Europe	United Kingdom	Barlow *et al.*, 2001	A, C, D, F, G, H	25
		Cane *et al.*, 2001	C	43
		Pillay *et al.*, 2002	A, AG, AE, C, D, F, G, H	67
		Frater *et al.*, 2001	A, C, D	18
	Spain	Perez-Alvarez *et al.*, 2001a,b	G, BG	31
Latin America	Brazil	Caride *et al.*, 2000, 2001	F, A	5
		Brindeiro *et al.*, 2002	A, C, F	17
	Cuba	Ruibal-Brunet *et al.*, 2001	A, C	16
M.East	Israel	Grossman *et al.*, 2001	B, C	73

Modified from Kantor and Katzenstein, 2004.

to PIs among HIV subtypes is also suspected to be an issue with subtype A and G viruses, due to the presence of multiple polymorphisms many of which are located near classically described PI mutations. For example, it has been shown that Nigerian HIV subtype A and G isolates frequently contain between nine and 17 amino acid substitutions within the protease gene, all of which would be classified as secondary mutations or polymorphisms (Agwale *et al.*, 2002). A comprehensive review of studies of drug resistance in non-subtype B HIV-1 is presented by Kantor and Katzenstein, 2004, summarised in Table 2.2.

It remains to be seen whether multi-drug resistant (MDR) viruses will emerge more rapidly during ART when such minor resistance conferring mutations are present before treatment initiation. As recent estimates state that in the United Kingdom over 50% of current infections are caused by non-subtype B viruses (Tatt *et al.*, 2004) it is likely to become increasingly important that we can accurately assign subtype to HIV-1 sequences.

2.1.3 Aims

Amino acid sequences contain less variation than nucleotide sequences due to the redundancy within the genetic code, and are not preferred for performing phylogenetic analyses. Handling amino acid sequence, however, significantly reduces the volume of sequence to be analysed and clinical management of sequences generated by drug

resistance testing often results in the storage of translated protein sequences rather than nucleotide sequence because the predictions of drug resistance are based on amino acids changes rather than nucleotide changes. Furthermore, in certain cases data is stored only as amino acid changes relative to a consensus sequence (subtype B). To subtype such a sequence the non-mutated sites must be derived from the consensus. Accordingly, to maximise the amount of data that we could analyse, the feasibility of subtyping protein sequences, namely a 430 amino acid region of the Pol polyprotein, generated as part of routine HIV-1 drug resistance monitoring (PR-RT), was investigated.

Specifically, different clustering methods were investigated, to see if established HIV-1 subtype phylogenies could be recreated using amino acid sequences. Then from initial subtype alignments for Gag, PR-RT and Env the relative levels of amino acid sequence variation between these proteins was determined, and for the PR-RT alignments position specific scoring matrices (PSSMs) that define the sequence variability of each subtype were created. These PSSMs are the basis of the subtyping tool, with which clinically derived PR-RT sequences may then be compared and subtype assigned. This method alleviates the need for extensive phylogenetic analysis and clustering of clinical sequences to accurately define a subtype, whilst utilising a method that uses rigorously characterised subtype profiles. Test clinical PR-RT sequence datasets were acquired from resistance testing facilities in London, Birmingham and the Caribbean.

The collection of subtype data at the population level and linking this to clinical data such as development of resistance, resistance mutations, time to AIDS and transmission may help reveal if any differences in these factors exist, between genetic variants of HIV. The development of the subtyping tool here described and its implementation in the National Resistance Database may go some way to initiating this kind of study.

2.2 Methods

2.2.1 Sequence Data

Full-length HIV-1 genomes and individual Gag, Pol and Env GenBank accession numbers were obtained from GenBank using scripts that were written to create and maintain the viral database VIDA (Alba *et al.*, 2001). Scripts that returned GenBank accession numbers of the viral genomes and proteins of the virus families within VIDA (modified by R. Myers) were used to search GenBank for HIV-1 genomes and proteins. Protein sequences corresponding to the accession numbers obtained from VIDA were downloaded from the National Centre for Biotechnology Information (NCBI http://www.ncbi.nlm.nih.gov/). 72 sequences used as HIV-1 reference subtypes were obtained from the Los Alamos HIV-1 Database (http://hiv-web.lanl.gov/content/index) (Table 1, Appendix).

2.2.2 Sequence Processing

In order to analyse the sequence variation within HIV-1 genomes, Gag, Pol and Env sequences from full-length HIV-1 genomes and Los Alamos reference sequences were combined to generate multiple alignments. This dataset was created by assembling a list of all complete HIV-1 genomes, together with their accession numbers for Gag, Pol and Env proteins. The three lists of accession numbers (Gag, Pol and Env) were compared with the NCBI non-redundant protein database (nr), downloaded as a FASTA formatted flatfile, and the HIV-1 sequences returned using a custom written Perl script. Each genome was then checked for the presence of Gag, Pol and Env proteins using a MS Access database, and in cases where there was a protein missing from a genome the relevant stretch of nucleotide sequence was translated using an in-house translation tool and added to the overall dataset. The accession numbers of HIV-1 subtype reference nucleotide sequences (Los Alamos Database) were incorporated into this database.

The final GENBANK and Los Alamos combined dataset included 170 Env proteins, 187 Gag proteins and 174 Pol proteins, the number of proteins is not equivalent as not all of the HIV-1 genomes were full length (~9 Kb), with proteins of partial size being excluded. The Pol dataset was further processed into a PR-RT dataset, the sequence length analogous to that produced for HIV resistance testing.

2.2.3 Phylogenetic Analysis

Three phylogenetic analyses were performed to group sequences by HIV-1 subtype, using the Los Alamos reference dataset to facilitate the partitioning of HIV sequences for the generation of sequence alignments that represent each HIV-1 subtype. These were maximum parsimony (Phylip, protpars), neighbour joining (ClustalW) and agglomerative hierachical clustering using complete linkage based on sequence identity matrices. The neighbour joining algorithm was implemented in ClustalW, after sequence alignment. To perform maximum parsimony, protein distances were calculated using Protdist and the distances applied to the maximum parsimony tree using Fitch, the tree generated using Protpars and Consense. All programs were part of the Phylip package. To perform the former analysis a programme for calculating sequence identity scores was used, taking it's input as a multiple sequence alignment and then returning a matrix of pairwise sequence identities based on the number of absolute amino acid matches along the length of two sequences from the original multiple alignment. The output was used to build linkage trees, using the software package STATISTICA (Version 6.1, StatSoft, Inc. USA). These methods were used to both curate subtype specific alignments for development of a subtyping algorithm, and to see if simple measures of genetic distance are sufficient to distinguish between different subtypes, using amino acid sequence, specifically in the PR-RT region. A table of all full-length sequences and their classification by each of the three methods is shown in Appendix, Table 2.

2.2.4 Sequence Variability

Subtype specific HIV-1 sequence alignments for Gag, Pol and Env were analysed to assess the levels of sequence variation that they contained. This was achieved using the in-house program conPSSM which, in addition to creating a consensus amino acid sequence for each subtype alignment, records the number of amino acid variants within a given alignment at each position in that sequence, compared to a consensus. These data enabled the construction of contour plots (using STATISTICA) which represent the deviation of individual sequences, away from a subtype B reference, and which is grouped by subtype so variation both within and between subtypes, across Gag, Pol and Env, may be analysed.

2.2.5 Mutation analysis

The frequency of resistance mutations present in subtype specific profiles and London-derived clinical data was assessed, using the custom written program Mutant. This compares each sequence to a consensus, and detects changes that have been characterised as either primary or secondary resistance mutations (Parikh *et al.*, 2001). The frequency of these in each subtype specific profile was calculated using MS Excel, expressed as the percentage of sequences in which each mutation was detected, in each profile.

2.2.6 PSSM generation and method validation

The subtype specific alignments generated for the PR-RT section of Pol were used for calculating subtype-specific PSSMs, which form the basis of the subtype analyser (STAR) subtyping tool. There were 11 subtype-specific multiple alignments for this genomic region, representing subtypes; A, AG, B, C, D, FK, G, H, J, N and O. As these multiple alignments were derived from one global multiple alignment they were in effect aligned with each other.

The process of PSSM building is implemented within the first stage of the STAR program. Specifically, the fraction of each amino acid at each position within each of the 11 subtype alignments was calculated and stored, giving a matrix (per subtype) where the first dimension was alignment position and the second dimension contained the fraction of each amino acid present at that position. By using the fraction of amino acids present at each position the matrices became normalised thereby removing the effect of variation in the number of sequence members present within each subtype alignment. Two extra elements were also added to the 20 amino acid elements in the second dimension, representing gaps and unassigned amino acids.

A method of resampling data in order to assess confidence in the ability to assign subtype (leave one out validation) was performed. This assessed the ability of STAR to accurately assign subtype, and also allowed the calculation of the mean and standard deviation of scores produced for the sequences that define a subtype PSSM, thereby representing the confidence thresholds for assigning sequences to subtype groups. The lower boundaries of the confidence threshold were calculated based on the average identity score for each subtype, minus 1 and 1.5 standard deviation measures.

2.2.7 Dataset updates

As GENBANK is continuously updated with new HIV sequence additions that are useful to expand under populated PSSMs, a rule base for updating subtype profile alignments was established to enable inclusion of the most recent PR-RT sequence data available from the most recent GenBank release (137 - August 2003). Using the BLAST algorithm GenBank NR was searched with HIV-1 PR-RT, returning approximately 28,000 sequences. These were sorted using a Perl script to retain only those sequences that contain both the protease and reverse transcriptase active site motifs DTG and YXDD, separated by 254 amino acids. 10,065 such sequences were aligned and all sequences containing greater than 2% ambiguous amino acid calls were removed, leaving 865 sequences. These 865 sequences were further filtered to remove sequences containing defined primary drug resistance mutations, with the exception of subtype O sequences known to have the natural polymorphism Y181C, otherwise associated with nevirapine exposure. NJ was used to subtype the resulting 654 sequences by clustering them with the sequences from original subtype profiles and by STAR. From these, only unambiguously subtyped sequences (372) were used to expand profiles to an arbitrary maximum of 100 sequences per PSSM.

2.2.8 Analysis of clinical datasets

Clinical sequences were obtained from the Health Protection Agency (HPA) laboratories in London (n=843) and Birmingham (n=885), where they were generated as part of standard resistance monitoring, and from the Caribbean Epidemiology Centre (CAREC, n=71) for subtyping using STAR.

2.3 Results

Currently the most extensive HIV sequence resources are those generated from clinical laboratories in developed countries, where routine genotypic drug sensitivity analysis is performed. It would be useful for subtype to be determined from such sequence, especially at source. As such data will continue to grow the aim of this chapter was to develop a subtyping tool for use in the clinical setting that will assign HIV-1 subtype based solely on the PR-RT amino acid sequence (named subtype analyser; STAR).

2.3.1 Creating sequence datasets

GenBank was searched extensively, in order to retrieve as many HIV-1 sequences as possible for the curation of subtype-specific sequence profiles, for the Gag, Pol and Env regions of the genome. The resulting dataset, including Los Alamos reference sequences, comprised 170 Env proteins, 187 Gag proteins and 174 Pol proteins - the number of proteins is not equivalent as not all of the HIV-1 genomes were full length (~9 Kb), with proteins of partial size being excluded. The 174 Pol sequences, (protease, reverse transcriptase, RNAse H and integrase) were processed further to give a set of proteins containing all of HIV-1 protease (99 amino acids) and the N-terminal 340 amino acids from reverse transcriptase (PR-RT). The combined 439 amino acid sequence is analogous to the maximum that is routinely sequenced as part of antiretroviral drug therapy monitoring. Four amino acid multiple sequence alignments (Gag, Pol, PR-RT, and Env) were generated using ClustalW, manually edited and used in subsequent phylogenetic analysis.

2.3.2 Phylogenetic analysis of HIV-1 Gag, Pol, PR-RT, and Env

Complex models of evolution are necessary for the most accurate phylogenetic methods of sequence analysis and, as such, these methods can be computationally very intense and time-consuming, especially when large numbers of sequences are processed. We therefore wished to determine whether a simple method of measuring distance between amino acid sequences would be sensitive enough to distinguish between HIV subtypes, along the length of the genome, compared to analysis using more complex phylogenetic methods. If so, then large numbers of non-subtyped sequences could be initially grouped by HIV-1 subtype using Los Alamos reference sequences within the dataset

(according to their placement within phylogenetic trees), and sequence alignments generated that represent expanded collections of sequences of defined HIV-1 subtype, for Gag, PR-RT and Env. Provided that there was sufficient divergence between subtypes in the PR-RT region, subtype-specific profiles could then be created which would contain as extensive a representation of the true diversity within each subtype in this genomic region as possible. This could then be used as the basis for a HIV-1 subtyping tool, which would take sequence data produced during resistance testing as its input.

Initially three phylogenetic analyses were performed on protein multiple sequence alignments, namely; maximum parsimony (MP, Phylip, protpars), neighbour joining (NJ, ClustalW) and agglomerative hierachical clustering using complete linkage based on sequence identity matrices (section 1.2.3). It was decided that where two of the three tree building methods agreed on the assignment of subtype to an unknown sequence, relative to a reference sequence, that sequence would be assigned as being of that subtype. The assignment of subtype was refined by generating additional trees (neighbour joining) from subsets of the Gag, Pol and Env alignments and checking the placement of sequences within those trees. In most cases a conservative strategy of removing sequences that did not position within trees consistently, was adopted. This resulted in the exclusion of sequences in an attempt to preserve the integrity of subtype alignments rather than trying to incorporate every possible sequence. The final dataset comprised 153 Gag sequences, 144 Pol (and therefore PR-RT) sequences and 140 Env sequences, classified into 11 subtype groups; A, AG, B, C, D, FK, G, H, J, N and O.

The final set of HIV-1 genomes containing Gag, PR-RT and Env proteins was given an in-house accession number of the form **00AABBCC_11** (Table 1/2, Appendix):
00 – corresponds to the number of the genome (numbers 1-72 Los Alamos sequences, numbers >500 – sequences derived from GenBank).
AA – Two-letter code used to represent the subtype of the Gag sequence, eg **AA** = subtype A, AB = subtype A / B recombinant, XX = unassigned.
BB – Two-letter code used to represent the subtype of the Pol sequence.
CC – Two-letter code used to represent the subtype of the Env sequence.
_11 – Refers to a key for country of origin.

Overall, the NJ and MP trees produced were remarkably similar. The trees generated by hierarchical clustering based on sequence identity matrices, however, were less similar and between all methods the placement of recombinant reference sequences was inconsistent.

2.3.2.1 Neighbour joining

Initially, NJ trees of Gag, PR-RT and Env amino acid sequence were created (Figure 2.2). The reference sequences within the trees are marked with coloured circles and in the case of all proteins studied these reference sequences cluster together in distinct branches, associated with non-reference sequences, presumably of the same subtype. The Gag and Env trees produce clusters with clear separation from each other, indicating that the increased variability in these regions of the genome enables easier separation of the different subtypes based on amino acid sequence. Despite the less 'neat' topology of the PR-RT tree, however, there data revealed that in the majority of cases each virus was classified as the same subtype in Gag, PR-RT and Env. To support the accuracy of this neighbour joining analysis, it was observed that recombinant reference sequences, for example CRF12-BF - a BF recombinant that is mainly subtype B within PR-RT, but F in Gag and Env - was classified as such (Figure 2.3). The same effect was observed for CRF01-AE, an AE recombinant that is subtype A within Gag and PR-RT, but E (a putative subtype) in Env. In this case, CRF01-AE Gag and PR-RT amino acid sequence clustered with A, but its Env sequence formed a distinct cluster (Figure 2.4). In some cases, recombinant non-reference sequences were detected: two sequences published as subtype C were classified as BC recombinants, across Gag, Pol and Env. These may be true recombinant genomes, or may represent discrepancies between the clusters generated for each gene.

Additional observations made from these trees are firstly, that the subtype F and K sequences cluster together in all cases, and are indistinguishable from each other in their branching pattern (Figure 2.2). If this observation holds true for trees built using other phylogenetic methods, these two subtypes should be merged to form a single profile. Secondly, CRF02-AG appears to cluster distinctly from subtypes A and G in all trees. This suggests that it may be possible to create a recombinant-specific profile for this genotype.

Figure 2.2 **Neighbour joining trees of Gag (A), PRRT (B) and Env (C) sequences.**
NJ was implemented in ClustalW to generate the trees shown.

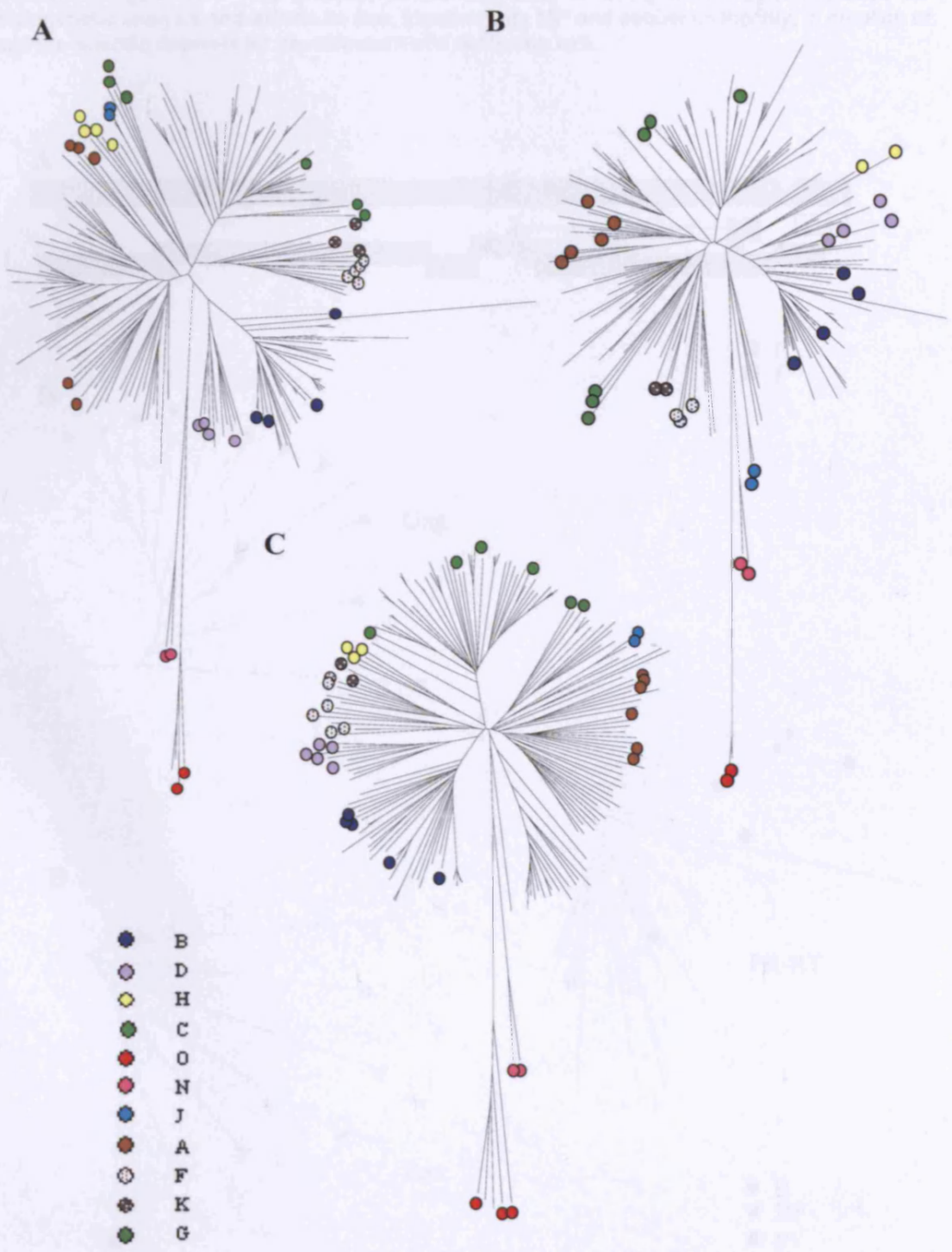


Figure 2.3 Example of identification of CRF012-BF, a BF recombinant, using NJ analysis of Gag, PR-RT and Env amino acid sequence.
The shaded bar **(A)** represents the genome structure, and **(B), (C)** and **(D)** show the subclusters within the Gag, PR-RT and Env NJ trees that reflect the recombinant structure of the BF recombinant genome. This analysis shows the utility of using amino acid sequence for phylogenetic analysis and affirms its use, together with MP and sequence identity, in creation of subtype-specific datasets for development of a subtyping tool.

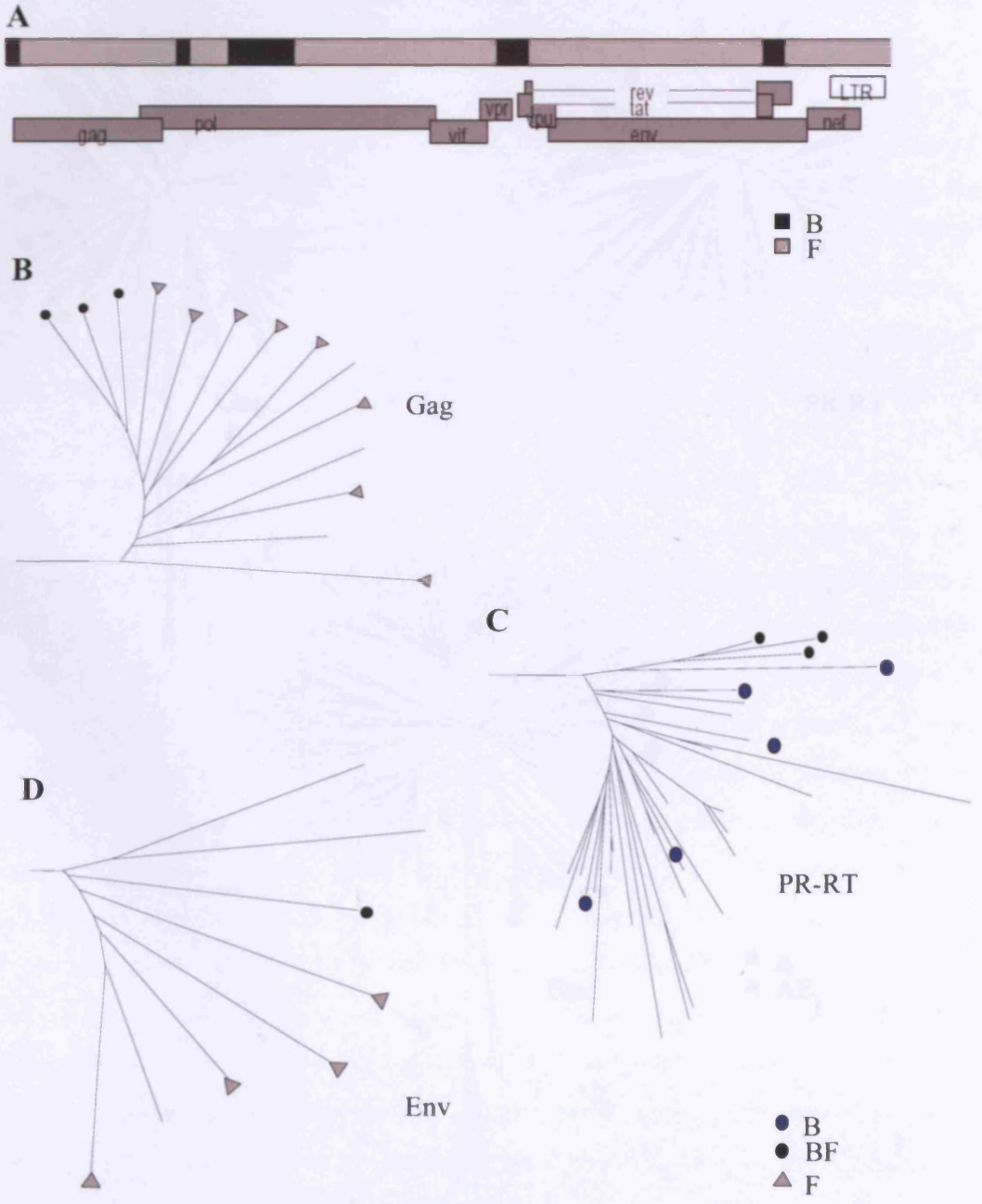
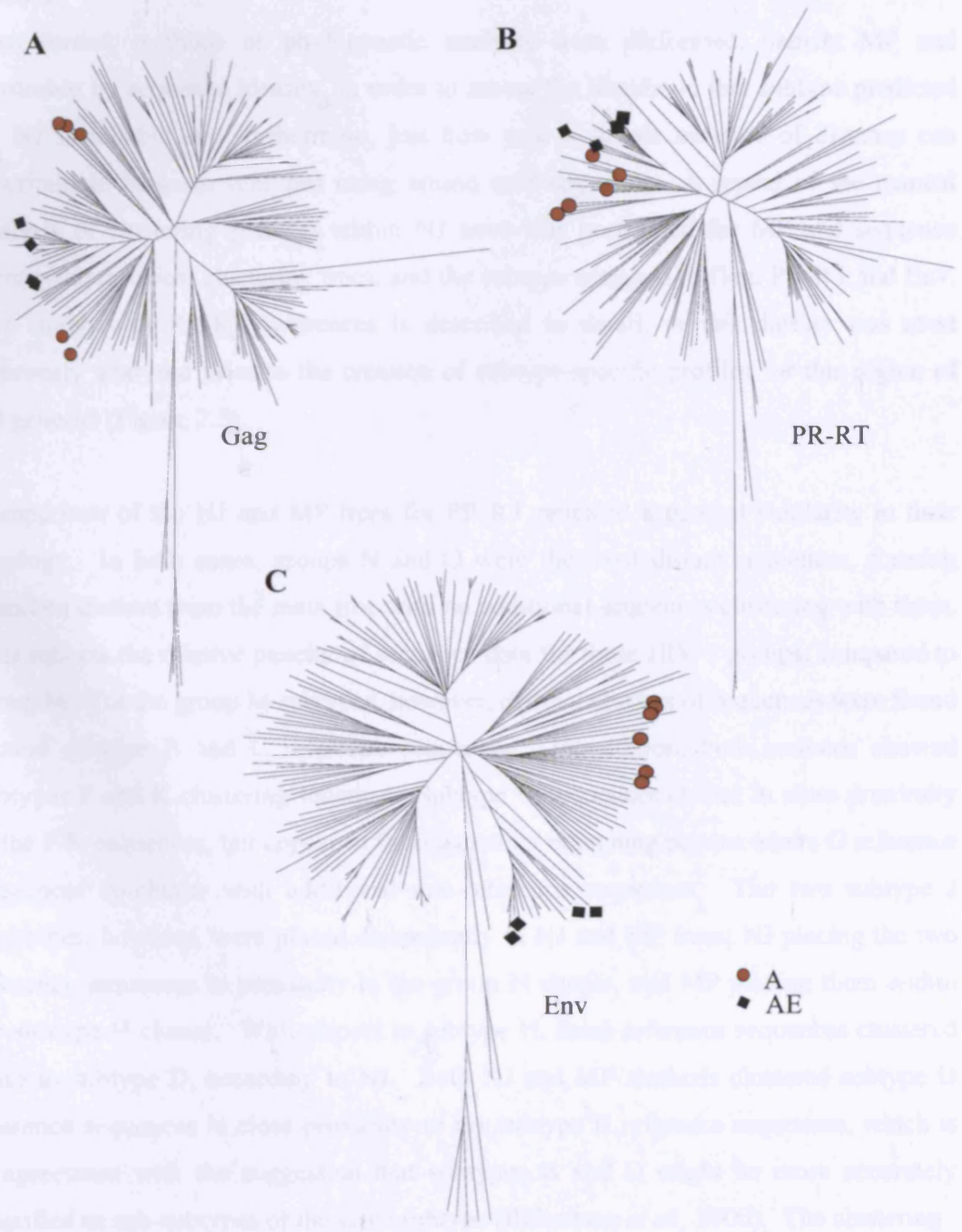


Figure 2.4 Example of identification of CRF01-AE, an AE recombinant, using NJ analysis of Gag, PR-RT and Env amino acid sequence.
The A and AE sequences cluster together within Gag and PR-RT (**A** and **B**), but separately within Env (**C**), indicating that the Env sequence is a subtype distinct from A, but not particularly closely associated with any other subtype.



From this initial analysis, use of amino acid sequence to subtype HIV-1 PR-RT appears sufficiently accurate. In order to test further the classifications made using NJ trees, therefore, MP and sequence identity trees were generated for Gag, PR-RT and Env.

2.3.2.2 Maximum parsimony and Hierarchical clustering based on sequence identity matrices

Two further methods of phylogenetic analysis were performed, namely MP and clustering by sequence identity, in order to assess the likelihood that subtype predicted by NJ is reliable and furthermore, just how well a simple measure of distance can discriminate between subtypes using amino acid sequence. A repeat of the manual analysis of clustering patterns within NJ trees was performed for MP and sequence identity/hierarchical clustering trees, and the subtype assigned to Gag, PR-RT and Env. The analysis of PR-RT sequences is described in detail, as this dataset was most rigorously analysed prior to the creation of subtype-specific profiles for this region of the genome (Figure 2.5).

Comparison of the NJ and MP trees for PR-RT revealed a general similarity in their topology. In both cases, groups N and O were the most distant sequences, forming branches distinct from the main tree with no additional sequences clustering with them. This reflects the relative paucity of sequence data for these HIV-1 groups, compared to group M. For the group M subtypes, however, distinct clusters of sequences were found around subtype B and C reference sequences. In addition, both methods showed subtypes F and K clustering together. Subtype G sequences cluster in close proximity to the F/K sequences, but appear to form a distinct branching pattern where G reference sequences co-cluster with additional non-reference sequences. The two subtype J sequences, however, were placed discordantly in NJ and MP trees; NJ placing the two reference sequences in proximity to the group N cluster, and MP placing them within the subtype H cluster. With respect to subtype H, these reference sequences clustered close to subtype D, according to NJ. Both NJ and MP methods clustered subtype D reference sequences in close proximity to the subtype B reference sequences, which is in agreement with the suggestion that subtypes B and D might be more accurately classified as sub-subtypes of the same subtype (Robertson *et al.*, 2000). The clustering

Figure 2.5 Maximum Parsimony and hierarchical clustering based on sequence identity trees of HIV-1 PR-RT sequences.
(A) The tree shown was generated using Protpars and Consense. Protein distances were calculated using Protdist and the distances applied to the maximum parsimony tree using Fitch. All programs were part of the Phylip package. **(B)** The tree distance matrix was generated using the in-house program SeqID and the tree constructed using Statistica (over page).

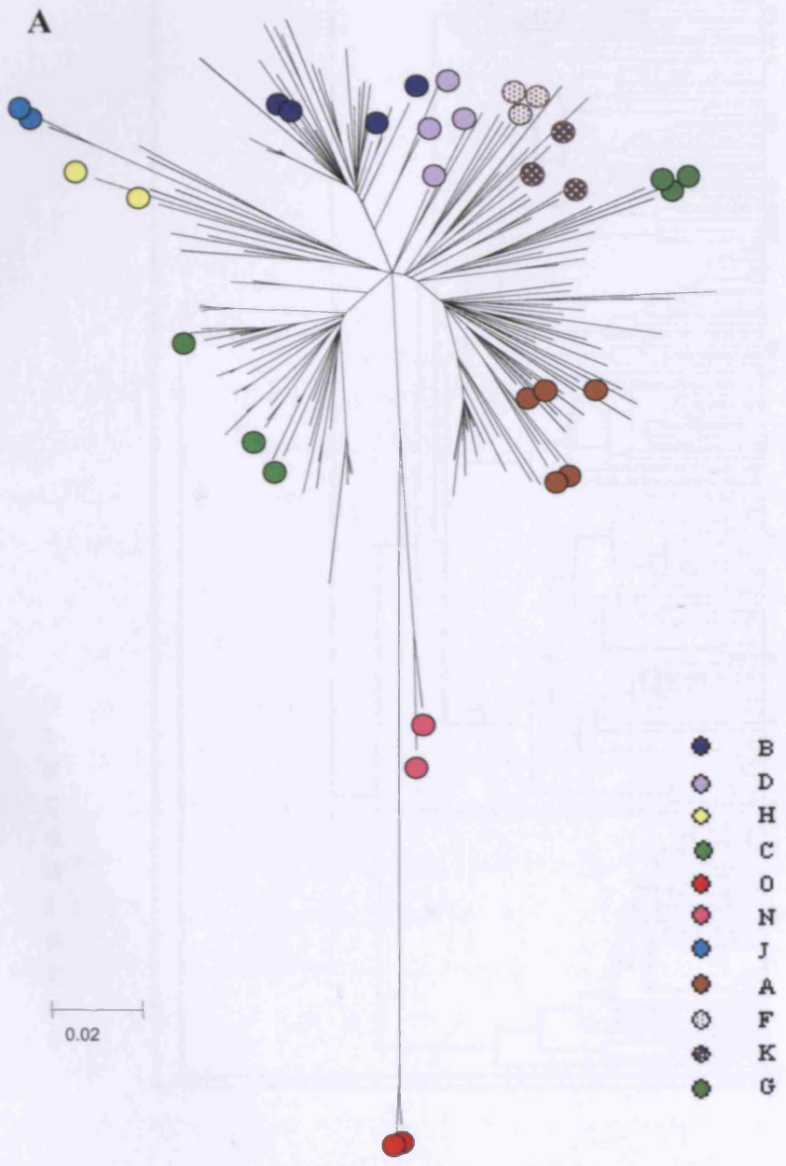
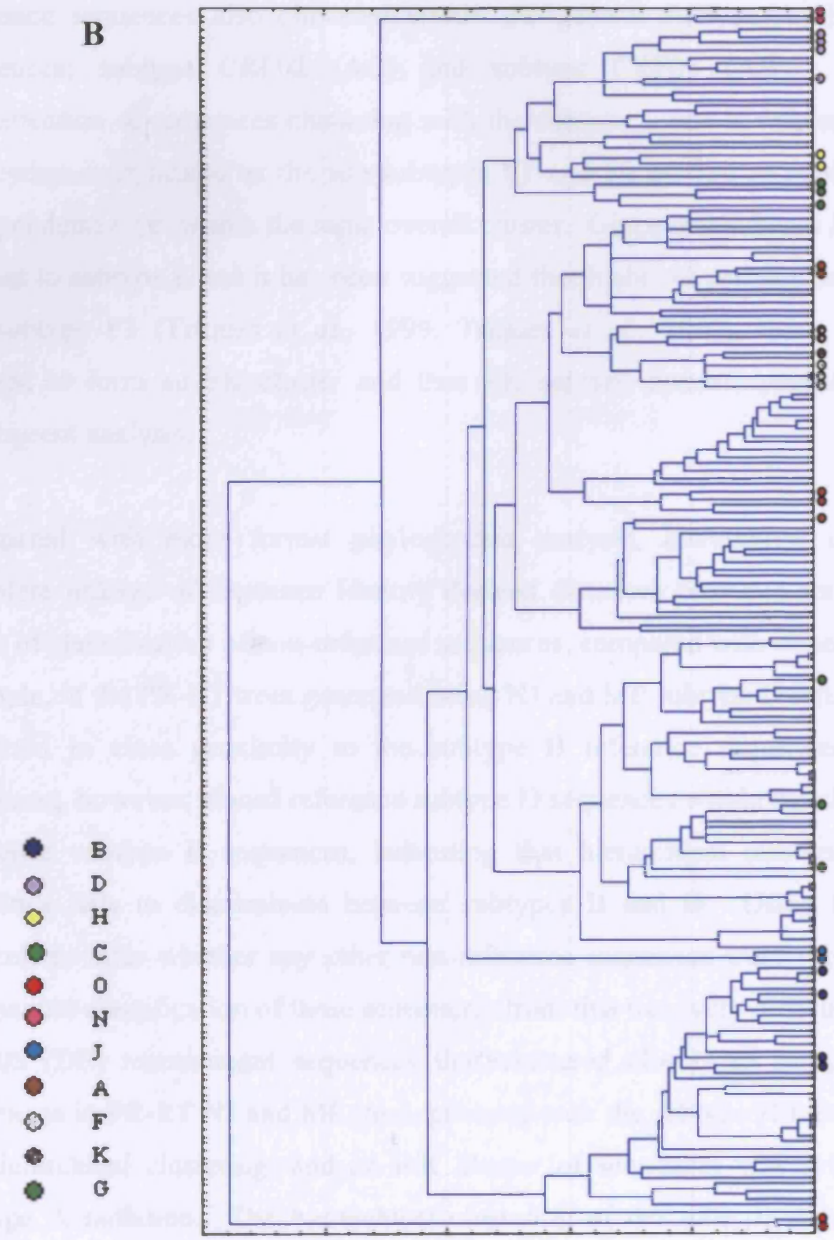


Figure 2.5 continued



of sequences around subtype A reference sequences within PR-RT showed that sequences of sub-subtypes A1 and A2 clustered separately, with sequences from the recombinant subtype CRF01 (AE) lying between them. Two other recombinant reference sequences also clustered within the general subtype A cluster of PR-RT sequences; subtype CRF02 (AG) and subtype CRF03 (AB). As mentioned, classification of sequences clustering with the subtype F and K reference sequences is somewhat complicated as the sub-subtypes F1 and F2 as well as subtype K clustered independently, yet within the same overall cluster. Given that subtype K is very closely related to subtype F and it has been suggested that it should perhaps be redesignated as sub-subtype F3 (Triques *et al.*, 1999, Triques *et al.*, 2000), these sequences were merged to form an FK cluster and thus FK subtype specific profile alignment, for subsequent analysis.

Compared with more formal phylogenetic analysis, hierarchical clustering using complete linkage of sequence identity derived distances provided some anomalies in terms of classification of non-reference sequences, compared with either NJ or MP. For example, in the PR-RT trees generated using NJ and MP subtype D reference sequences clustered in close proximity to the subtype B reference sequences. Hierarchical clustering, however, placed reference subtype D sequences within the cluster containing reference subtype B sequences, indicating that hierarchical clustering of sequence identities fails to discriminate between subtypes B and D. Using this tree, it was difficult to infer whether any other non-reference sequences were subtype D, making the manual classification of these sequences, from this tree, very difficult. Furthermore, CRF05 (DF) recombinant sequences that clustered closely to subtype D reference sequences in PR-RT NJ and MP trees, grouped with the subtype H reference sequences by hierarchical clustering, and an F/K cluster of sequences was evident within the subtype A radiation. The hierarchical clustering of the subtype A reference PR-RT sequence, however, was in good agreement with the classifications made using NJ and MP. CRF01 (AE) reference sequences clustered between sub-subtype A1 and A2 sequences and CRF02 (AG) clustered independently but close to the subtype A reference sequence cluster.

Thus, whilst some were in agreement, certain sequence classifications made using hierarchical clustering based on sequence identity were contrary to those made using

more complex phylogenetic methods. It is perhaps the tree building stage of classification using sequence identity (using hierarchical clustering in STATISTICA) that limits this method in its use as a phylogenetic tool, rather than the measure of sequence identity itself. A summary of the level of agreement between subtyping methods is presented in Table 2.3. Although NJ and MP are more consistent, hierarchical clustering performs well enough to suggest that simple measures of distance are sufficient to allow subtype assignation in PR-RT. Given its agreement with MP, the use of NJ to classify any further sequence data into subtype-specific datasets was deemed appropriate.

Table 2.3 Agreement between phylogenetic methods in the subtype classification of PR-RT amino acid sequences from non-classified HIV-1 sequences from complete genomes, in GENBANK.

No. Seq	169
% Total Agreement	94.08
% NJ/D	94.67
% NJ/MP	96.45
% D/MP	95.27

NJ – neighbour joining
MP – maximum parsimony
D – Sequence identity / hierarchical clustering

2.3.2.3 Compilation of subtype alignments

Following manual assignation of subtype to each sequence in the PR-RT dataset by three methods and derivation of a consensus, alignments were generated for 11 HIV-1 subtypes: 9 single subtypes; A, B, C, D, G, H, J, N and O, one composite; FK, as subtypes F and K were difficult to separate on phylogenetic trees, and one recombinant; AG (Table 2.4). CRF02 (AG) is a circulating recombinant form of subtypes A and G, which was well populated in terms of sequence numbers and formed clusters independent of subtype A across Gag, PR-RT and Env sequences. CRF01 (AE), however, was not separated into a discreet alignment for this analysis, as it is subtype A in PR-RT. Thus, the AE sequences were added to the A PR-RT alignment, explaining the increased number of subtype A Gag and PR-RT sequences, compared to Env. This variation in the number of sequences attributed to each region of the genome, for each subtype, is largely a product of recombinant sequences being identified. Furthermore, certain sequences were removed from the analysis where it was not immediately clear to

Table 2.4 **Distribution of Gag, Pol/PR-RT and Env sequences assigned to subtypes.**

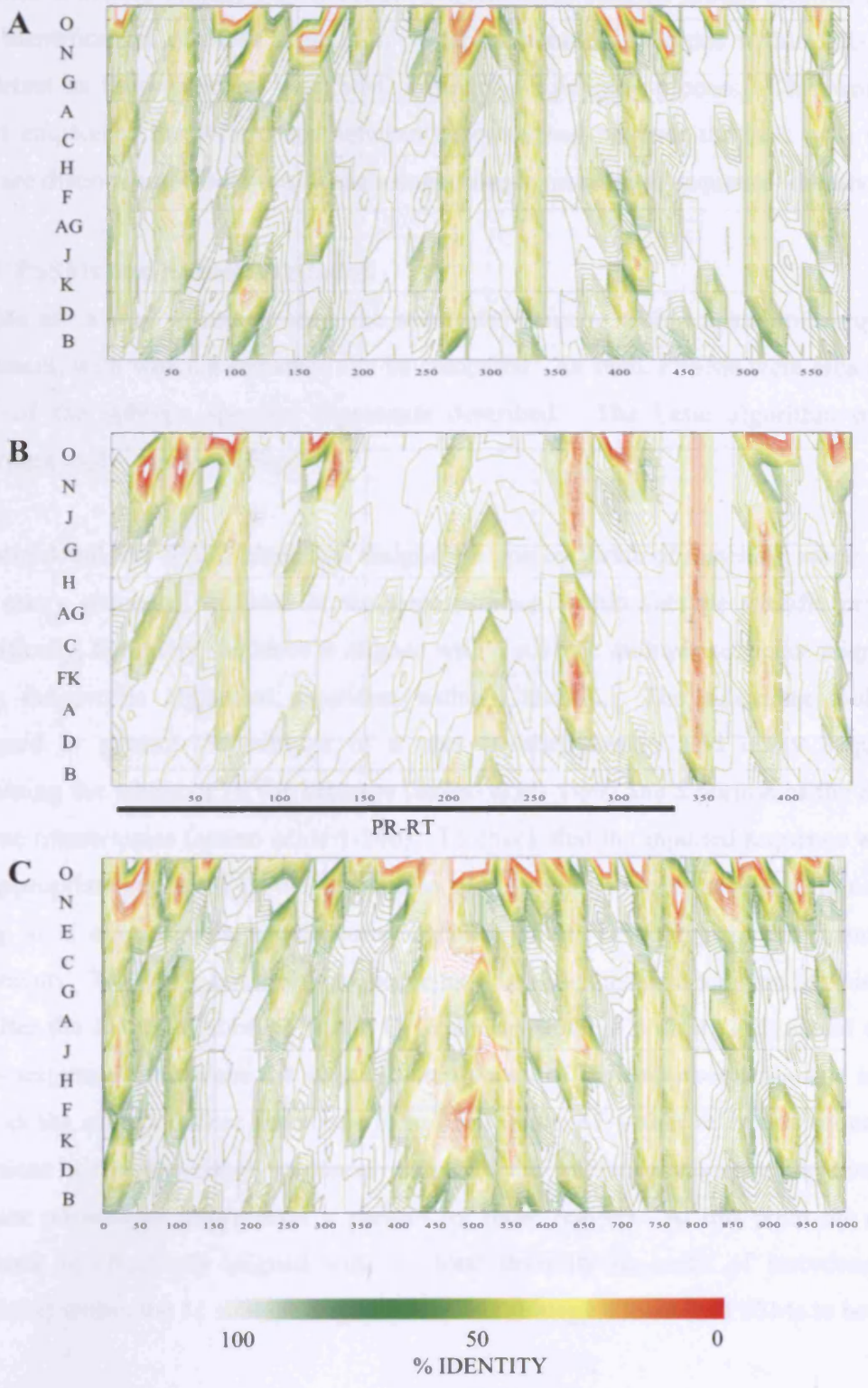
Subtype	Gag	Pol/PR-RT	Env
A	28	26	20
AG	10	16	13
B	36	33	33
C	44	39	41
D	8	8	7
FK	10	9	11
G	4	4	5
H	4	3	3
J	3	2	2
N	2	2	2
O	4	2	3
Total	153	144	140

which subtype cluster they belonged. Whilst this rationalisation of data reduces the number of sequences within each subtype-specific alignment, it removes the detrimental effect of using sequences that are difficult to classify and thus may compromise the efficiency of a subtyping tool, based on such data.

2.3.3 Variation in HIV across genome – suitability for subtyping?

Several reports have described that Pol sequence variation is sufficient to distinguish between subtypes (section 2.1.2), and phylogenetic classification of sequences based on PR-RT sequence seems robust. To better characterise the sequence variation across the HIV genome in the data collected and classified for STAR, subtype specific sequence variation was therefore analysed along the length of the Gag, Pol (including PR-RT) and Env proteins. A consensus sequence generated for the subtype B alignment was compared to each of the subtype specific alignments established previously and the percentage of disagreement at each position was used to generate a trace for each subtype that shows sequence differences relative to subtype B, displayed as a contour map (Figure 2.6). Different subtypes were ordered relative to their similarity to the subtype B consensus sequence; subtype B is at the bottom of each graph and subtype O at the top, with a general trend of increasing variation between the two. Traces derived from both Gag (Figure 2.6-A) and Env (Figure 2.6-C) proteins show constant variation along the length of the proteins, with many regions showing intense variation, such as amino acids 530 to 570 in Env. This corresponds to stem IIIB, IIB', IIC, IIC', IIA, III-IV and III-IV' of the rev response element (HIV Sequence Compendium, 2001). Such

Figure 2.6 Protein sequence variation in (A) Gag (n = 153), (B) Pol (n = 144) and (C) Env (n = 140) relative to a consensus reference B subtype sequence.
Regions of sequence variability are shown using contours with the degree of sequence identity illustrated by the colour change; green (high) through yellow to red (low).



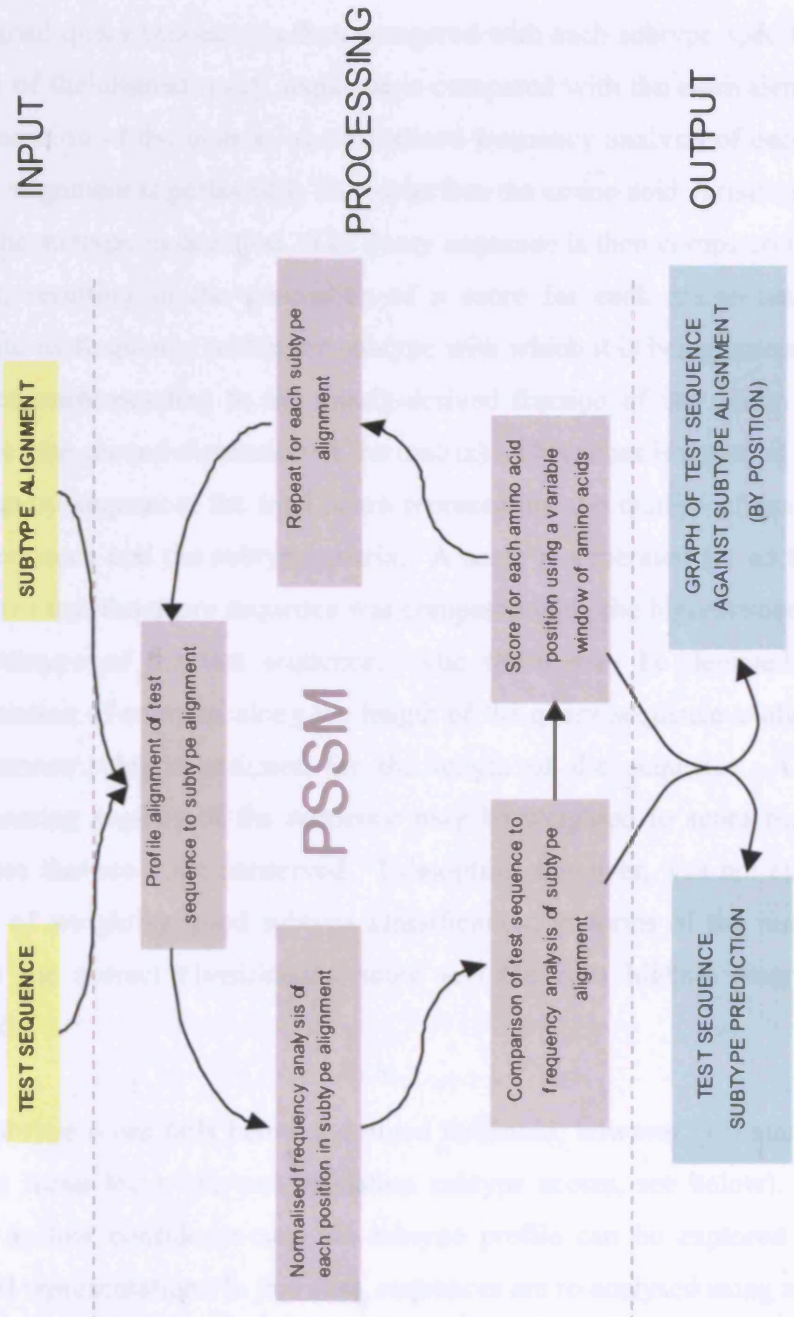
highly variable regions were generally not subtype specific, but represented regions of higher variability in all subtypes. Although less variable, however, subtype specific 'hotspots' of variation exist in the PR-RT region of Pol (Figure 2.6-B), and it is probably these subtype discriminating regions that enabled the NJ and MP phylogenetic methods to clearly classify HIV-1 PR-RT sequence into subtype groups (section 2.3.2). The identification of these regions of variation between subtypes within PR-RT is important as they could be 'weighted' during the subtyping process. This would in effect emphasise the differences between subtypes and increase the ease with which they are discriminated between, when using a simple measure of sequence identity.

2.3.4 PSSMs and method validation

PSSMs are a way of representing the sequence variation within a multiple sequence alignment, with which a sequence can be compared. As such, PSSMs were created for each of the subtype specific alignments described. The basic algorithm of this subtyping tool is shown in Figure 2.7.

In more detail, the STAR algorithm assigns subtype by virtue of matching amino acids in a query sequence, to those at the same position within subtype specific profiles. Specifically, the query sequence is aligned with a subtype multiple sequence alignment using the profile alignment algorithm within ClustalX. The subtyping tool was designed to predict the subtype of a user inputted amino acid query sequence, containing the whole of HIV-1 protease (amino acids 1-99) and a portion of the HIV-1 reverse transcriptase (amino acids 1-340). To check that the inputted sequence was of the appropriate region of HIV-1 Pol and to trim or extend the ends of the amino acid string, it is aligned with a consensus sequence of all 11 subtype specific multiple alignments. Whilst the consensus sequence used is biased towards subtype M, this does not alter the ability of subtype N and O sequences to align with it. Alignment of the query sequence determines the start and stop points of the query sequence and inserts gaps in the regions where insertions have been observed within reference sequences. Insertions within the query sequence relative to the reference sequences are excised, because no subtype information is present for those regions. At this point the query sequence is effectively aligned with the total diversity (in-terms of insertions and deletions) within the 11 subtype alignments. This strategy allows the PSSMs to be

Figure 2.7 Schematic of the subtyping tool, name subtype analyser: STAR.



calculated once only at the start of the process (before any sequences are inputted), rather than having to recalculate the subtype PSSMs with each new query sequence.

The aligned query sequence is then compared with each subtype-specific matrix. Each position of the aligned query sequence is compared with the equivalent position in the first dimension of the matrix. A normalised frequency analysis of each position in the subtype alignment is performed. This describes the amino acid variation at each position within the subtype in question. The query sequence is then compared to this frequency analysis, resulting in the generation of a score for each amino acid in the query, indicating its frequency within the subtype with which it is being compared. A score is generated corresponding to the matrix-derived fraction of the amino acid in the test sequence (the second dimension of the matrix). This score is summed along the length of the query sequence, the total score representing the quality of match between the query sequence and the subtype matrix. A score is generated for each of the subtype alignments that the query sequence was compared with, the highest score suggesting the likely subtype of the test sequence. The result may be depicted as a graphical representation of subtype, along the length of the query sequence analysed, or returned as consensus subtype assigned for the length of the sequence. Certain subtype-discriminating regions of the sequence may be weighted to score more significantly than those that are more conserved. This option, however, was not explored as in the absence of weighting good subtype classification, in terms of the numerical distance between the correct classification score and the next highest incorrect score, was observed.

If the subtype score falls below a defined threshold, however (1.5 standard deviations from the mean leave one out validation subtype scores, see below), the sequence is flagged as low confidence and the subtype profile can be explored further through graphical representation. In this case, sequences are re-analysed using a sliding window method. Each position in the query sequence is considered together with 14 flanking amino acids, giving a default window size of 15. These 15 positions in the query sequence are compared to all aligned subtype profiles. The average frequency score across the window, per subtype, is then assigned to the amino acid at the centre. The window is then moved forwards in increments of one amino acid, until each position in the sequence has been scored. This option enables the detection of recombinants in the

PR-RT region, provided that there is sufficient genetic distinction between recombining HIV-1 subtypes.

Prior to its use with clinically derived input sequences, however, an assessment of the ability of this method to accurately classify a set of test sequences of known subtype was performed. The sequence dataset as a whole did not contain sufficient numbers of sequences to allow partitioning into a training and test dataset, so a leave one out analysis strategy was adopted in order to evaluate the performance of STAR. This involved removing each sequence from its subtype alignment, recompiling the subtype PSSM in its absence, and testing the ability of the method to correctly classify the sequence when removed from the subtype-specific PSSM. From an initial dataset of 174 sequences, 144 were used in this analysis as the remaining 30 sequences were classified as recombinant forms and hence were not included in the subtype alignments.

STAR correctly predicted all but one of the sequences during the leave one out analysis, predicting one subtype H sequence as subtype D (Table 2.5). This was an improvement over the initial implementation of STAR (Gale *et al.*, 2004), where 5 misclassifications were made. This improvement was achieved by the amalgamation of subtype F and K datasets, as suggested by NJ and MP trees (section 2.3.2).

Table 2.5 **Leave One Out Analysis of STAR subtyping tool.**

Subtype	n.	Pos	Fp	Fn
A	26	26	0	0
AG	16	16	0	0
B	33	33	0	0
C	39	39	0	0
D	8	8	1	0
FK	9	9	0	0
G	4	4	0	0
H	3	2	0	1
J	2	2	0	0
N	2	2	0	0
O	2	2	0	0
Total	144	143	1	1

Accuracy	0.99
Coverage	0.99
Precession	0.99

n. – number of sequences in subtype
Pos – number of sequences correctly identified
Fp – number of sequences incorrectly assigned to a subtype
Fn – number of sequences that were not assigned to a subgroup
Accuracy = pos/n.
Coverage = pos/(pos+fn)
Precision = pos/(pos+fp)

2.3.4.1 Updating PSSMs

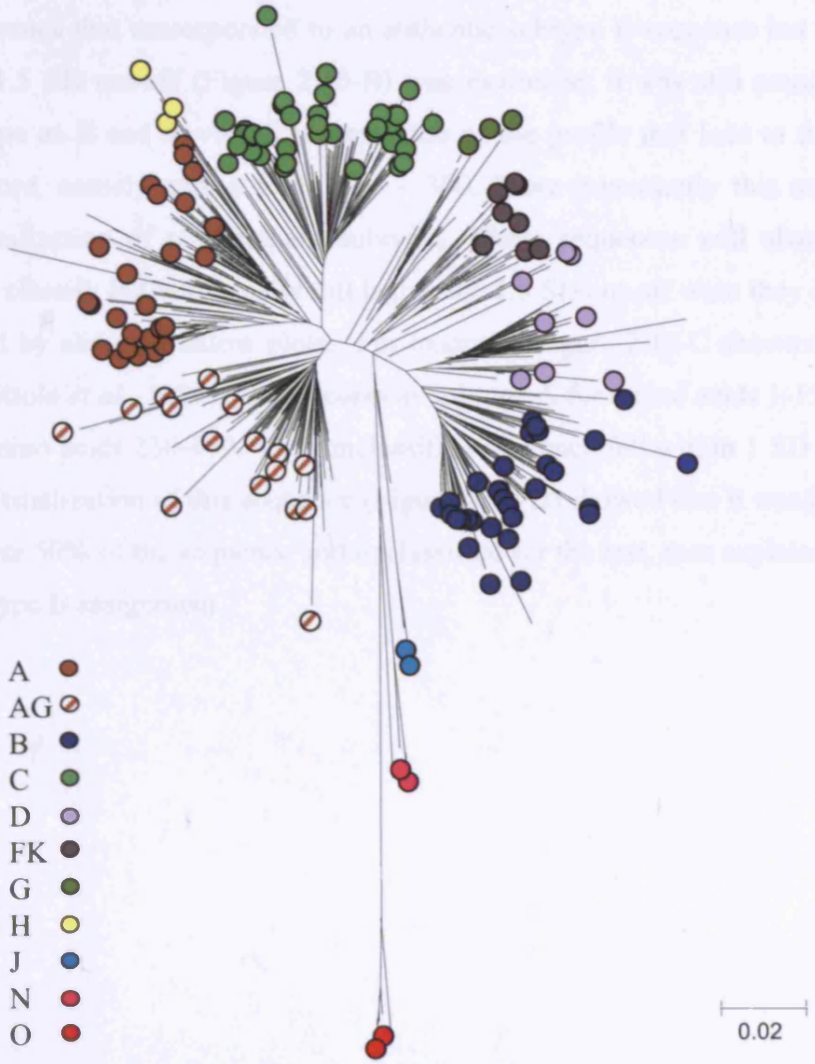
Each Genbank release provides more sequences with which to populate PSSMs. Not all sequences are of appropriate quality to be used (i.e. abundant ambiguity codes), so a rule base for updating PSSMs was devised (Gale *et al.,* 2004). This identified 372 candidate sequences, which were analysed relative to the core 144 sequences, by building a neighbour joining tree (Figure 2.8). These subtyped sequences were used to further populate each subtype profile to an arbitrary maximum of 100 sequences. Specifically, A, AG, B and C profiles each contain 100 sequences, D, 50; F/K, 30; G, 18; H, 4; J, 3; N, 5; 0, 6. Following the update of subtype profiles, leave one out validation confirmed that inclusion of new sequences in profile alignments did not significantly affect performance of the algorithm. Three of 516 sequences were misclassified, giving an accuracy rate of 98.8%. These 3 sequences were incorrectly assigned as subtype AG instead of subtype A (2) and subtype C (1).

2.3.5 Recombinants and confidence thresholds

Whilst leave one out analysis was used to assess the ability of the subtype analysis tool to accurately assign subtype, as described it also allowed the calculation of the mean and standard deviation of scores produced for the sequences that define a subtype PSSM, thereby representing the confidence thresholds for assigning sequences to subtype groups.

Specifically, based on the classification of the 516 defined subtype sequences (this time including known recombinant sequences, previously removed – section 2.3.4), a lower threshold score for confidence in subtype prediction could be assigned. Two lower thresholds were assessed, namely 1 and 1.5 standard deviations (SD) below the mean subtype score for the resampled leave one out validation data. At the more stringent 1 SD an average of 14% of subtype assigned sequences fell below the threshold and were therefore flagged as being of putative subtype. This number decreased to 7.7% at 1.5 SD from the mean, the distribution being equivalent among all subtypes. A typical example of score distribution is shown for the resampled subtype B alignment (Figure 2.9) where 5 sequences from 100 are flagged at 1.5 SD below the mean (5%). Analysis of 11 sequences identified as recombinants containing subtype B sequence within the Pol region (Los Alamos Database) showed only 3 sequences (BF recombinants) scored

Figure 2.8 **Neighbour joining tree of HIV-1 PR-RT amino acid sequence for dataset updating purposes.**
Tree generated from 516 sequences, 144 used within the initial STAR program and 372 additional sequences derived from GenBank. The positions of the initial 144 sequences classified into 11 subtype groups are indicated with coloured dots.



within 1.5 SD of the subtype B alignment. The remaining recombinant sequences (8) were all flagged as being of putative subtype.

For such 'flagged' sequences, by virtue of the sliding window scoring function, a graphical representation of the subtype along the length of the query sequence can be extracted to aid subtype prediction (Figure 2.10). As an example, four sequences from subtype B were visualized. Subtypes classified as within 1.5 SD of the mean subtype score are well demarcated from the next most similar high scoring subtype (Figure 2.10-A). When a sequence that corresponded to an authentic subtype B sequence but which fell bellow the 1.5 SD cut-off (Figure 2.10-B) was examined, it was still possible to assign the subtype as B and to visualize the region of the profile that lead to the low classification score, namely amino acids 190 – 300. More importantly this analysis enables the visualization of recombinant subtypes. These sequences will always be more difficult to classify but will usually fall below the 1.5 SD cut-off were they can be readily identified by sliding window plots. For example, Figure 2.10-C shows an AB recombinant (Liitsola *et al.*, 1998) which scores as subtype A for amino acids 1-151 and subtype B for amino acids 230-499. One unclassified sequence fell within 1 SD of the subtype mean. Visualization of this sequence (Figure 2.10-D) showed that it was indeed subtype B for over 50% of the sequence and unclassified for the rest, thus explaining its low scoring subtype B assignment.

Figure 2.9 Distribution of STAR scores within the subtype B leave one out analysis.
Dots (•) are used to indicate non-recombinant subtype B sequences. Squares (□) indicate recombinant sequences containing some subtype B sequence. The 4 sequences indicated by broken lines are plotted in figure 2.10. **(A)** (Fig 2.10-A) illustrates a high scoring subtype B sequence. **(D)** (Fig 2.10-D) shows a high scoring recombinant BF sequence. The three recombinant sequences (□) scoring above the lower 1.5 standard deviation threshold are all BF recombinants. **(C)** (Fig 2.10-C) represents an AB recombinant scoring below the lower threshold. **(B)** (Fig 2.10-B) highlights one of the 2 non-recombinant subtype B sequences that score below the lower threshold.

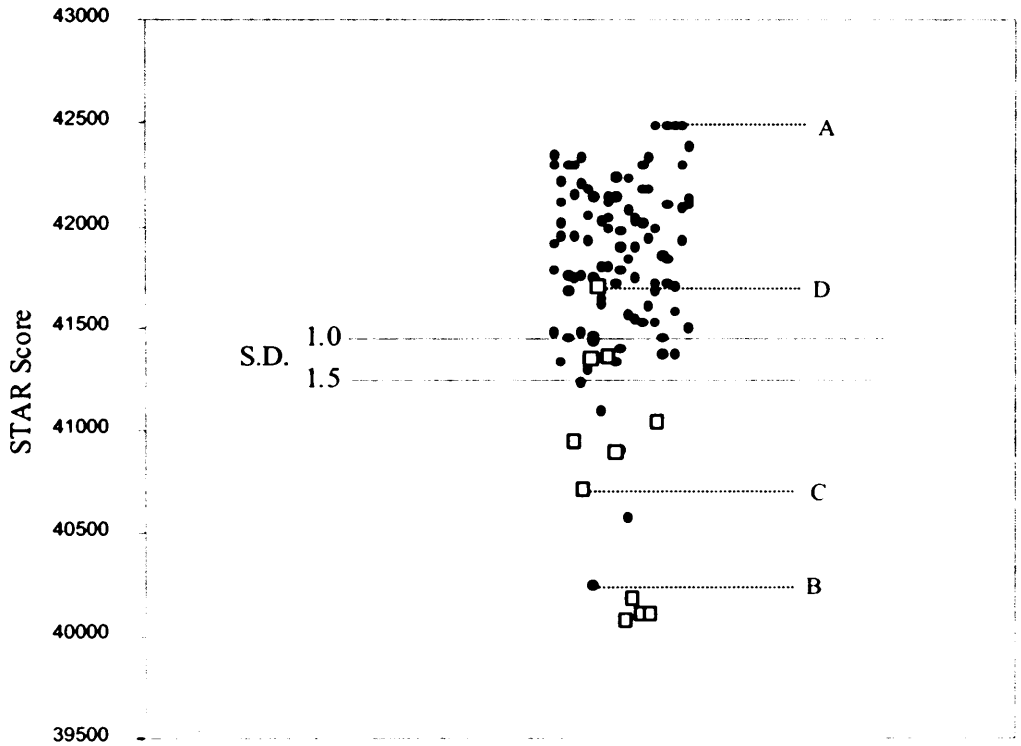
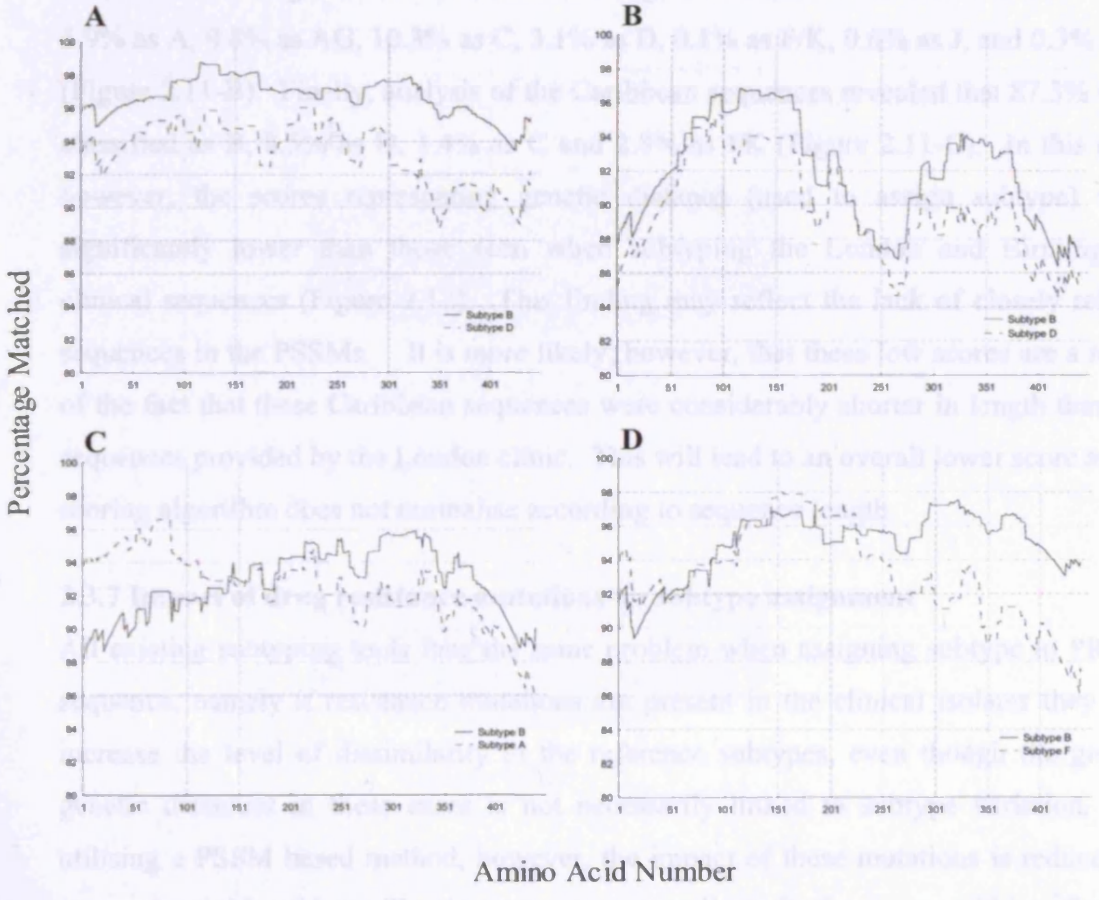


Figure 2.10 Graphical visualisation of the 4 sequences highlighted in Figure 2.9.
Each graph plots amino acid position against the percentage match to a sliding window of amino acids along the sequence length. The solid line traces the score against the subtype B PSSM and the dashed line traces the score against the second highest scoring PSSM. **(A)** and **(B)** show non-recombinant subtype B sequences, with **(A)** showing a sequence that generates a high subtype B STAR score (GenBank accession AF256209) and **(B)** shows a low confidence "flagged" subtype B sequence with low scoring regions against the subtype B PSSM making it poorly separated in some regions from the next highest scoring PSSM (GenBank accession U26546). **(C)** and **(D)** show recombinant sequences. **(C)** shows an AB recombinant sequence that would be detected as recombinant with a high score against subtype A, before switching to a high score against subtype B. This sequence (GenBank accession AF193277) has been identified previously as an AB recombinant. **(D)** shows a plot of one of the 3 high scoring BF recombinants. The BF recombinant (GenBank accession AF385935) scores similarly to subtypes B and F initially and then scores more highly compared with subtype B along the last 200 amino acids.



2.3.6 Testing the method – clinical dataset

Given the accuracy of the new subtyping algorithm, as defined by leave one out validation, PR-RT sequences were obtained from the resistance screening facility, Department of Virology, UCL (n=843), the resistance screening facility, Health Protection Agency (Birmingham, n=885) and from the Caribbean Epidemiology Centre (CAREC, n=71). Nucleotide sequences were translated into amino acid sequence where necessary and the sequences were subtyped as a batch, by submitting a FASTA formatted multiple sequence file to STAR. Subtyping was achieved at a rate of 36 sequences per minute (Intel Pentium IV processor with 2.5 Ghz RAM). Results indicate that, of 843 PR-RT sequences from the London clinic, STAR classified 84% with high confidence: 72.1% as B, 10.5% as A, 4.4% as AG, 5.7% as C, 6.1% as D, 0.7% as F/K and 0.6% as G (Figure 2.11-A). In the Birmingham dataset 78.9% were classified as B, 5.9% as A, 0.8% as AG, 10.3% as C, 3.1% as D, 0.1% as F/K, 0.6% as J, and 0.3% as G (Figure 2.11-B). Finally, analysis of the Caribbean sequences revealed that 87.3% were classified as B, 8.5% as D, 1.4% as C and 2.8% as FK (Figure 2.11-C). In this case, however, the scores representing genetic distance (used to assign subtype) were significantly lower than those seen when subtyping the London and Birmingham clinical sequences (Figure 2.12). This finding may reflect the lack of closely related sequences in the PSSMs. It is more likely, however, that these low scores are a result of the fact that these Caribbean sequences were considerably shorter in length than the sequences provided by the London clinic. This will lead to an overall lower score as the scoring algorithm does not normalise according to sequence length.

2.3.7 Impact of drug resistance mutations on subtype assignment

All existing subtyping tools face the same problem when assigning subtype to PR-RT sequence, namely if resistance mutations are present in the clinical isolates they will increase the level of dissimilarity to the reference subtypes, even though the greater genetic distances in these cases is not necessarily linked to subtype variation. By utilising a PSSM based method, however, the impact of these mutations is reduced as any amino acid residue will only score proportionally to its frequency within reference datasets. Provided resistance mutations are not present in subtype profiles, they will be excluded from the subtyping process. Accordingly, the frequency of such mutations in both reference sequences and clinical isolates of defined subtype (London dataset) were

Figure 2.11 Pie charts showing the percentages of each subtype identified within the London (A), Birmingham (B) and Caribbean (C), datasets, as classified by Star.

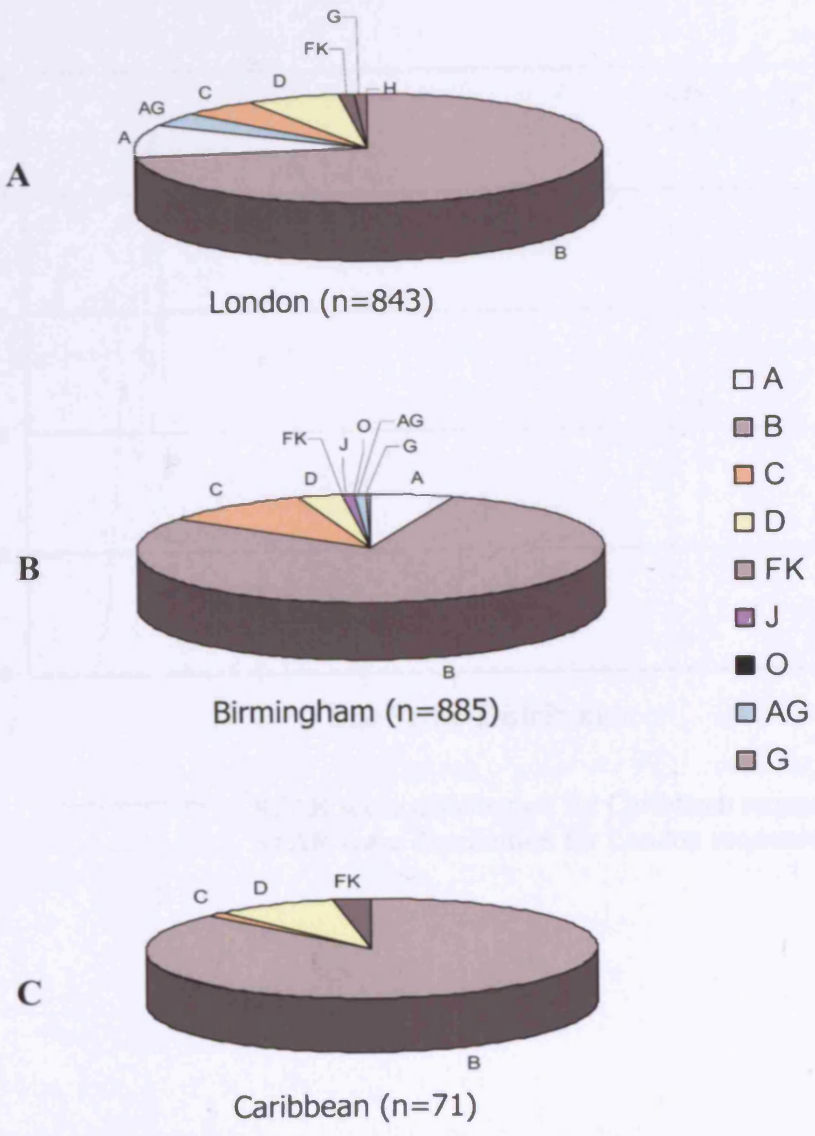
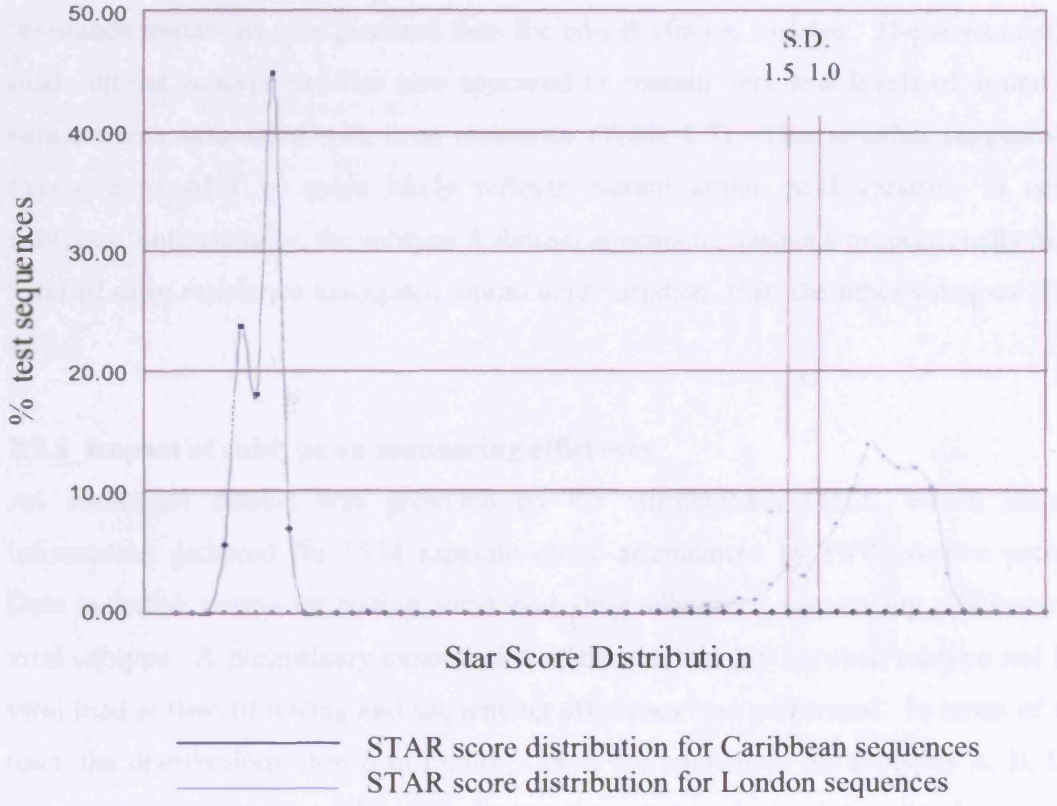


Figure 2.12 Distribution of Star scores for the Caribbean and London test datasets. This figure shows the considerably lower STAR scores for Caribbean PR-RT sequences, compared to London sequences. The scores are below the confidence threshold assigned as a result of leave one out analysis, used to identify poor sequences and potential recombinants in the test and London datasets.



determined. Specifically, the in-house program Mutant was used to report all mutations within a query sequence that are not due to natural variation and select those sequences containing key mutations that can be considered to indicate a history of exposure to antiretroviral therapy (Parikh *et al.*, 2001, Table 2.6). The frequency of such mutations in subtype clusters of clinical and reference sequences are shown in Figure 2.13. Sequences from the London clinic showed a consistently high frequency of resistance mutations in subtype B sequences with approximately twice the average number of resistance mutations (per genome) than the non-B clinical isolates. The sequences that made up the subtype profiles also appeared to contain very low levels of amino acid substitutions associated with drug resistance (Table 2.7). This is either suggestive of exposure to ART or more likely reflects natural amino acid variation in certain subtypes. Interestingly, the subtype A dataset appears to contain a proportionally higher level of drug resistance associated amino acid variation, than the other subtypes (Table 2.7).

2.3.8 Impact of subtype on sequencing efficiency

An additional dataset was provided by the Birmingham PHLS, which included information gathered for 1534 separate clinic attendances by HIV positive patients. Data included; reason for testing, viral load, drug adherence, sequencing efficiency and viral subtype. A preliminary examination of the relationship between subtype and both viral load at time of testing and sequencing efficiency was performed. In terms of viral load, the distributions shown in Figure 2.14-A are equivalent for subtypes A, B, C, D and G, suggesting that within this dataset there is no clear relationship between the subtype of virus with which patients are infected, and their viral load at time of testing. Interestingly, however, the data presented in Figure 2.14-B shows a clear relationship between sequence failure and HIV-subtype. For patients where sequencing and subtype classification has been achieved on at least one previous occasion, subsequent sequencing failure for subtype B viruses occurs in only 5% of cases, whereas for all other non-B subtypes sequencing failure rates are substantially higher. Whilst for subtypes AE (n=11), AG (n=2), G (n=9) and K (n=2) these results could be a product of the very small sample sizes studied, for subtypes A (n=97), C (n=166) and D (n=54) the sequencing failure rates of 20, 26 and 14% respectively seem to be a clearly significant observation in relation to resistance sequencing methodologies, if not statistically.

Table 2.6 **Resistance mutations (primary and associated) used to classify sequences as drug exposed and/or potentially drug resistant (number represents amino acid position).**

Protease	
Primary	30, 32, 48, 50, 82, 84, 88, 90
Associated	46, 47, 54, 71
Reverse Transcriptase	
Primary	41, 65, 69, 70, 74, 75, 100, 101, 103, 115, 181, 184, 190, 215, 219
Associated	118, 210

Figure 2.13 **Resistance mutations within different subtypes.**
Mean number of resistance mutations within PR and RT, from HIV-1 clinical isolates (classified by STAR, from the London dataset) and present in original reference sequences comprising the subtype-specific PSSMs (obtained from GENBANK and Los Alamos).

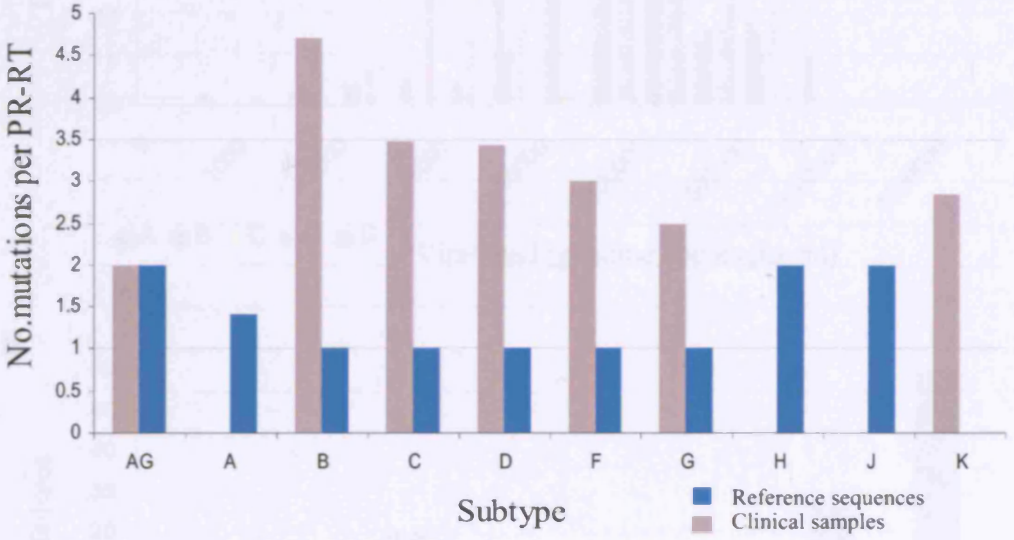
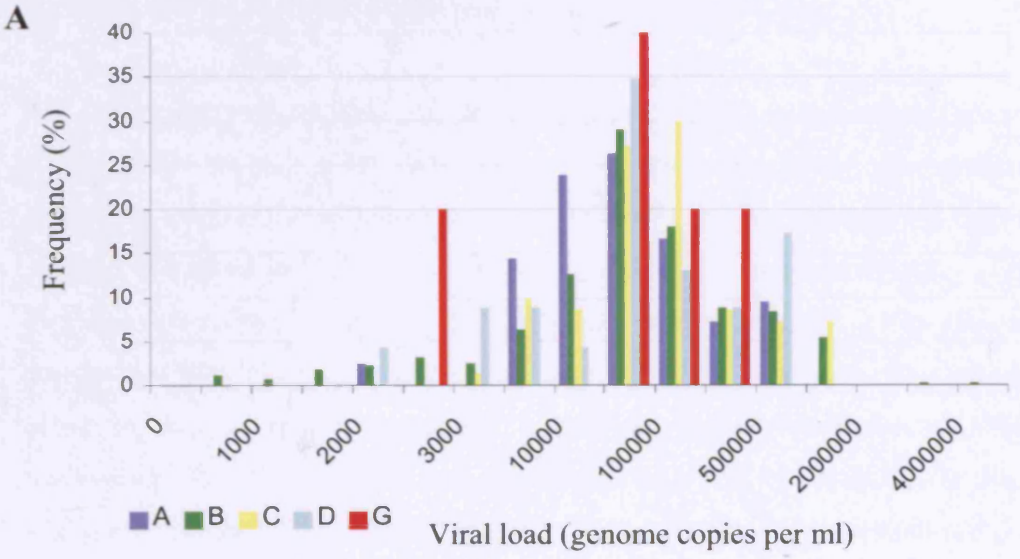
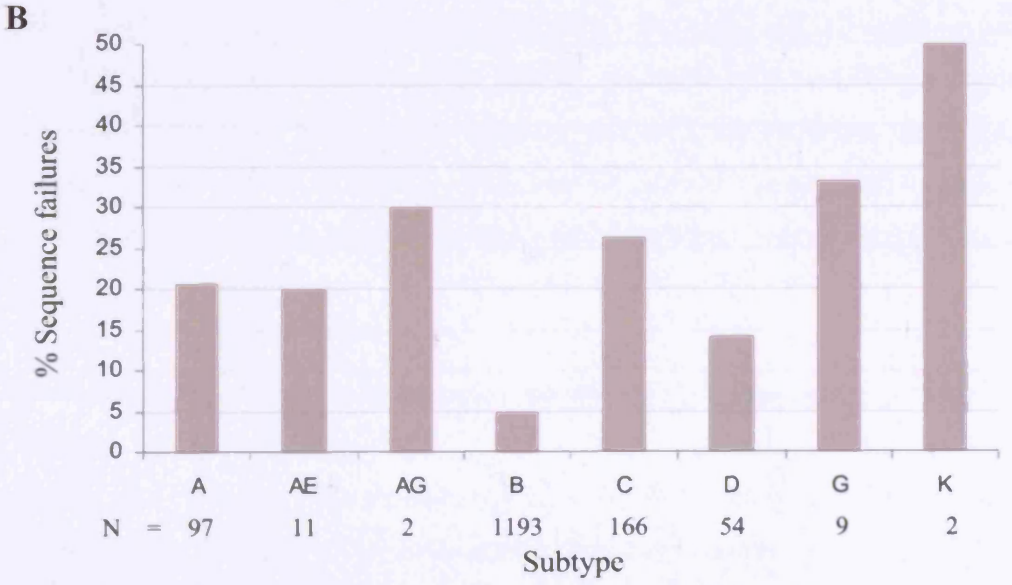


Table 2.7 **Percent prevalence of resistance mutations in subtype reference datasets (pre-update).**

	Subtype								
RT mutations	*A*	*AG*	*B*	*C*	*D*	*F*	*G*	*H*	*J*
K219Q	***7.7***	***12.5***	0	0	0	0	0	***33.3***	0
K65E/R	***3.9***	***6.3***	0	0	0	0	0	0	0
K70R	***19.2***	***12.5***	***3***	0	0	0	***25***	***33.3***	0
L100S	0	0	0	***2.6***	0	0	0	0	0
M41L	0	0	0	0	***12.5***	0	0	0	***50***
T215I/Y/S	***15.4***	0	***6***	0	0	0	0	0	***50***
T69N/D/I/S	***7.7***	***6.3***	0	***2.6***	0	***14.3***	0	0	0
PR mutations	*A*	*AG*	*B*	*C*	*D*	*F*	*G*	*H*	*J*
A71T	***3.9***	0	0	0	0	0	0	0	0
D30N	0	0	0	***5.1***	0	0	0	0	0
M46I	***100***	0	0	0	0	0	0	0	0
No. sequences	*26*	*16*	*33*	*39*	*8*	*7*	*4*	*3*	*2*

Figure 2.14 Relationship between viral load, sequence failure and subtype.
(A) Viral load distribution in subtype A, B, C, D and G infections, based on data from 1519 individual testing occasions. **(B)** Percentage of sequence failures of all reported resistance sequencing of PR-RT, for HIV-1 subtypes A, AE, AG, B, C, D, G and K. The number of reports in each category (N) is noted, of which the bar represents percentage failure.





In terms of sequencing efficiency at least, this data reveals that patients infected with non-subtype B HIV-1 are not being managed optimally with respect to monitoring drug resistance. This therefore highlights the importance, in a population where 21.1% of patients harbour non-subtype B viruses (Figure 2.11-B), of determining viral subtype as soon as infection is diagnosed. Only if this is known can patient genotypic monitoring strategies be optimised.

2.4 Discussion

As one of the most extensive sources of HIV sequence data, clinical resistance testing provides a highly valuable data resource. If such sequence data were to be accurately classified at source, in terms of subtype, the resulting information would prove invaluable in the study of the distribution and clinical significance of the various HIV-1 subtypes, and aid in the optimisation of patient management strategies.

The most accurate method of grouping amino acid sequences is by extensive phylogenetic analysis. Methods such as maximum likelihood, however, are user dependent, computationally expensive and unsuitable for clustering large datasets such as those produced in clinics carrying out HIV genotyping. Accordingly, a subtyping tool named Subtype Analyser (STAR) has been developed here, which allows comparison of a query sequence against the full genetic diversity of a characterised group of subtype-specific sequences, which are defined as tables of amino acid frequencies, based on multiple sequence alignments. STAR intentionally uses amino acid sequences as clinical HIV-1 resistance sequences are often reported and databased as amino acid changes relative to a reference sequence. STAR can subtype such data following reconstruction of the full-length amino acid sequence. In addition, because of the high nucleotide variation rate of HIV-1, synonymously variable positions within nucleotide sequences may rapidly saturate and as such confound nucleotide based subtype classification schemes. The use of shorter amino acid sequences also considerably increases the computational speed with which subtype classification can be achieved.

2.4.1 There is sufficient sequence divergence in Pol to distinguish between subtypes

As Gag and Env have been traditionally used for HIV subtype classification, following curation of amino acid sequence datasets representing subtypes A, B, C, D, FK, G, H, J, group N and group O for Gag, Pol, PR-RT and Env proteins, using three separate phylogenetic methods, an investigation of the sequence variability between all the subtypes for which profiles were created, across Gag, Pol and Env, was performed. This revealed that the least sequence variability was seen in the Pol region (Figure 2.6), as expected. Subtypes B and D showed the greatest similarity, suggesting that it may be

difficult to distinguish between these two subtypes using a subtyping tool based on amino acid sequence. It is clearly very important, however, to be able to distinguish between these two genetic variants of HIV. Subtype D viruses are typically sub-Saharan African in origin, whereas subtype B predominates in Western Europe and the USA. In terms of monitoring the spread of non-B subtypes into these regions therefore, subtypes B and D must be readily identifiable based on PR-RT if this subtyping algorithm is to be a useful epidemiological tool. Phylogenetic analysis, coupled with leave one out validation studies showing a mean 99% accuracy, however, suggest STAR is both robust and applicable to subtyping HIV-1 based on the PR-RT regions sequenced for routine drug sensitivity testing. This agrees with previously published data (Pasquier *et al.*, 2001, Njouom *et al.*, 2003). The variation within amino acids sequences from HIV-1 PR-RT, when assessed in the context of STAR, are sufficient to allow discrimination between the genetically most similar, as well as more distantly related subtypes.

The creation of tables of amino acid frequencies based on multiple sequence alignments (subtype-specific profiles) required the careful curation of subtype-specific PR-RT sequence datasets. The subtype specific alignments generated by the initial phylogenetic analyses and sequence analysis were accordingly refined, including the merging of CRF01 (AE) sequences with the subtype A profile, as this CRF is subtype A within PR-RT. A distinct CRF02 (AG) profile, however, was created as representatives of this CRF clustered separately from subtypes A and G. Subtypes F and K were merged because they were difficult to distinguish between, perhaps unsurprising as it has been proposed that subtype K should be reclassified as sub-subtype F3 (Triques *et al.*, 1999, Triques *et al.*, 2000, Van der Auwera *et al.*, 2001).

2.4.2 Different patterns of resistance mutations were detected in subtype profiles

Drug resistance mutation analysis of sequences within subtype-specific profiles was performed, as the absence of resistance mutations from profiles was a prerequisite for them not influencing the accuracy of subtyping, especially of drug-exposed viruses. Analysis of the original reference sequences showed a very low level of resistance mutations, as required. Certain sequences, however, were found to possess certain resistance-associated mutations presumably present as natural polymorphisms. Subtype A sequences in particular appear to have a proportionally higher level of drug resistance

associated amino acid polymorphisms, compared to the other subtypes (Table 2.7). If this variation is reflected in all subtype A viruses it is possible that they may progress to ART resistance quicker or via different pathways than other genotypes, a suggestion which warrants further investigation. Such observations have been made for subtypes O and J previously (Descamps *et al.*, 1997) highlighting the potential importance of subtype classification in assessing and perhaps predicting treatment outcomes. At present, however, HIV subtype is an uncertain factor when making clinical decisions about antiretroviral therapy. Whilst current research may soon help to elucidate the important subtype-dependent differences that affect virologic response to therapy, it is hoped that by providing a reliable method of HIV subtyping this may be made easier and may help drive the expansion of drug treatment efforts, while simultaneously contributing to their effectiveness across different geographical regions and HIV subtypes. It is also interesting that a difference in the efficiency of resistance sequencing was detected, for different HIV-1 subtypes (Figure 2.14-B): sequence failures were proportionally much higher for non-B subtypes. It may therefore not only be treatment strategies, but also monitoring and diagnostics, which require adjusting when a particular patient's subtype status is known.

2.4.3 Testing of a clinical dataset and inclusion of a measure of confidence

As part of the STAR algorithm, a basic means of assessing the reliability of each subtype classification was created: if the classification score falls below a defined score threshold, it is marked as low confidence ('putative subtype') and may be further examined by virtue of a sliding window option (section 2.3.4). This option enables the detection of recombinants in the PR-RT region, provided that there is sufficient genetic distinction between recombining HIV-1 subtypes (Figure 2.10). The inclusion of a method of assessing the reliability of a subtype assignation is novel, and was included as other subtyping tools currently available do not provide this level of rigour.

To demonstrate the utility of STAR, 843 HIV-1 PR-RT sequences derived from clinical samples, from London; 885 HIV-1 PR-RT sequences derived from clinical samples, from Birmingham; and 71 HIV-1 PR-RT sequences derived from clinical samples, from the Caribbean, were subtyped using this new tool. The distribution of HIV-1 subtypes within this dataset reflects the previously reported prevalence of non-B strains in the UK (Parry *et al.*, 2001), and in the Caribbean (Vaughn *et al.*, 2004). Furthermore, 84%

of the subtype assignations made for UK-based patients are made with high confidence, a measure not available with any other web-based batch subtyping tool. This level of confidence is a virtue of the fact that the STAR algorithm has been designed to take into account the inherent variability within sequences produced for resistance evaluation. That is, sequence features such as resistance mutations (identified in the clinical test dataset) are corrected for within the scoring system due to their extremely low frequency of occurrence in subtype profiles. Such heterogeneity in clinical isolates increases the level of dissimilarity to the reference subtypes, resulting in greater genetic distances not necessarily linked to subtype variation. With STAR these factors do not affect the subtyping process.

Interestingly, however, whilst the Caribbean sequences were classified with high accuracy, the actual scores resulting in classification of each sequence were lower than that seen for the sequences classified from UK clinics (Figure 2.12). As a result, all sequences fell below the 1.5 SD confidence level, indicating that sequences which are from locations geographically distinct from those of sequences within the subtype profiles (used to set the confidence threshold), and which are produced by an alternative resistance testing protocol resulting in generation of a shorter PR-RT sequence, are less likely to score with high confidence. Because of this observation alternative scoring methodologies for STAR are in development which will overcome the sensitivity of this current measure of confidence in classification, both to more rare sequences not represented within subtype profiles and to shorter Pol sequences.

2.4.4 Recombinants within PR-RT may be identified

The occurrence of novel circulating recombinant forms (CRFs) of HIV-1 are being increasingly reported. CRFs, however, can only be detected by sub-genomic subtyping tools if the signature recombination site(s) are within the region used for classification. For example, STAR cannot identify CRF01 which is subtype A in the 5' region of the genome and subtype E in the 3' region. Accordingly, the PR-RT sequence of this AE CRF clusters with subtype A sequence. In the case of CRF02 (AG), however, this recombinant strain possesses sufficient genetic divergence in PR-RT to distinguish it from other subtypes. This enabled the creation of a recombinant subtype profile for the PR-RT region; hence STAR can identify CRF02 without the need for extra analysis.

Furthermore, by creating a confidence threshold and assigning 'putative subtype' flags to sequences that score below 1.5 SD of the subtype mean score it is possible to identify further recombinants in PR-RT (Figures 2.9 and 2.10). Thus, STAR may detect AG recombinants outright, or other types of recombination in the PR-RT region provided there is sufficient genetic divergence between the recombining subtypes. It should also be noted that the methodology behind STAR is not confined to analysis of the PR-RT region of HIV-1. If other regions of the HIV-1 genome become routinely sequenced for drug resistance, STAR could be expanded to encompass their classification.

Chapter 3.0 Phenotypic characterisation of HIV-1 subtypes and HIV-2 *in vitro*

3.1 Introduction

As described in the introduction to this thesis, several distinct subtypes and an increasing number of recombinant forms are responsible for the current HIV-1 pandemic. Despite the involvement of multiple genetic variants of HIV in the causation of disease worldwide, however, there has been a large skew in research focus on subtype B, which is the most prevalent subtype in the western world. As a consequence, it is currently not clear whether genetically distinct HIV-1 subtypes have different biological properties that cause differences in disease progression, or transmissibility. The accurate subtype classification of HIV isolates, for example using STAR, and the linking of this information to clinical observations, at a population level, may go some way helping answer these questions.

On the whole, however, it remains to be seen whether the properties and consequences of HIV infection can be generalised across the non-B subtypes that affect the majority of infected persons in the developing world. A limited number of preliminary studies have attempted to address the question as to whether HIV-1 subtypes differ; at the population level in their transmission and disease progression, or *in vitro* in terms of their fitness, replication capacity and sensitivity to antiretroviral drugs.

3.1.1 Epidemiological studies

The rate of HIV-1 disease progression varies widely between infected individuals and it was demonstrated as early as 1989 that variability in the biological properties of HIV isolates is one of the factors influencing the course of HIV infection (Tersmette *et al.*, 1989). A significant correlation was found between the mean replication rate of isolates obtained from an individual and the rate of CD4+ T-cell decline observed: in individuals with low-replicating HIV isolates no significant CD4+ T-cell loss was observed, but recovery of high-replicating isolates, in particular syncitium inducing (SI) isolates, was associated with rapid decline of CD4+ T-cell numbers and development of

AIDS. With the subsequent association of the SI/NSI characteristics of HIV with sequence variation in the *env* gene (Cheng-Mayer *et al.*, 1990) and the discovery of the main HIV-1 co-receptors CCR5 (Deng *et al.*, 1996, Dragic *et al.*, 1996, Alkhatib *et al.*, 1996) and CXCR4 (Feng *et al.*, 1996) the link between coreceptor usage and disease progression was established (Connor *et al.*, 1997). Several reports have also linked sequence variation among divergent HIV-1 subtypes with differing coreceptor usage, *in vitro* cytopathogenicity and therefore potentially disease progression (de Wolf *et al.*, 1994, Rubsamen-Waigmann *et al.*, 1994, Zhong *et al.*, 1995, Tscherning *et al.*, 1998). In terms of host effects, as described in Chapter 1.0, a deletion allele of the CCR5 co-receptor gene has been reported to slow disease progression in heterozygotes compared with individuals homozygous for the normal CCR5 gene (Dean *et al.*, 1996). Age at seroconversion has also been shown to influence disease progression (Pezzotti *et al.*, 1996). Other human allelic variants shown to influence the susceptibility to HIV-1 infection and/or the subsequent rates of disease progression include: genes that control viral entry into susceptible cells (chemokines and chemokine receptors); genes involved in immune regulation, such as interleukin-10 (IL-10), interleukin-4 (IL-4), tumour necrosis factor-alpha (TNF-alpha), and mannose-binding lectin (MBL); and genes involved in the adaptive immune recognition by T cells, (human leukocyte antigen (HLA) type; reviewed by Anastassopoulou and Kostrikis, 2003, Kaslow *et al.*, 2005). Both host and virus therefore play important roles in directing the progression of disease over time. It could be argued, however, that host effects are more likely to induce differences in disease progression between infected individuals, as the host has a great capacity for behavioural and external changes (such as nutrition or coinfection) that can be shown to directly influence the progression of disease. This is compounded in that, besides replication rate and coreceptor usage, few other virus-specific characteristics have been directly attributed to particular differences in disease phenotype. Epidemiological studies relating virus type to disease progression, however, have provided some evidence that suggests HIV genotype may play an important role in HIV disease.

3.1.1.1 Disease progression

Early studies of different subtypes of HIV-1 and their association with disease progression provided somewhat conflicting data. It was initially shown that the rates of

disease progression, evaluated by the decline in CD4 lymphocyte count, were similar between groups of Ethiopian immigrants infected with HIV-1 subtype C and non-Ethiopian Israelis infected with subtype B (Galai *et al.,* 1997). This study, however, compared two groups receiving antiretroviral therapy with different demographic composition, mode of transmission and host genetics. In contrast, a study from Senegal reported that individuals infected with HIV subtype A were found to progress clinically more slowly than those infected with non-A subtypes (Kanki *et al.,* 1999). In this study women infected with a non-A subtype were eight times more likely to develop AIDS than those infected with subtype A. However, another study investigating the influence of genetic subtype of HIV-1 and the ethnic origin of the infected individual on the rate of disease progression found no significant differences in the rate of disease progression or CD4 lymphocyte decline between individuals infected with subtypes A, B, C or D (Alaeus *et al.,* 1999). Furthermore, a study in Thailand reported that HIV-1 subtypes B and E (AE) were associated with similar degrees of immunosuppression and opportunistic infection patterns among a group of HIV-1 infected individuals (Amornkul *et al.,* 1999), but in this case all individuals were AIDS patients. Finally, in London it was reported that differences in CD4 lymphocyte count decline and progression to AIDS in HIV-1 infected Africans and non-Africans could not be attributed to ethnicity or viral subtypes (Del Amo *et al.,* 1998). In this study, however, it was presumed that the Africans were infected with subtypes A and C, and non-Africans with subtype B, making such data very hard to interpret.

The studies described, therefore, used cohorts composed of individuals who may or may not have been receiving antiretroviral therapy, were not homogenous in terms of host genetic composition, mode of transmission, demographic composition, or time of infection (Galai *et al.,* 1996, Del Amo *et al.,* 1998, Alaeus *et al.,* 1999) and largely generated conflicting results. Accordingly, subsequent epidemiological studies looking at the relationship between HIV-1 subtype and disease progression have typically aimed to investigate populations where these problems are controlled. One such study looked at 164 individuals infected with envelope subtype A or D viruses in a Ugandan population, one hundred and seventeen of whom had estimated dates of seroconversion, HIV-1 infection being almost exclusively acquired by heterosexual transmission and where antiretroviral drugs were not available (Kaleebu *et al.,* 2001). It was reported that, whilst disease progression in participants infected with subtypes A and D was not

significantly different for most end-points, there was a tendency for those infected with subtype D to progress faster than those infected with subtype A: there were over three times more deaths with subtype D compared to subtype A. This disparity between subtypes A and D, however, did not reach statistical significance for most indicators of progression. This work was followed by a second, larger study, in which the effect of HIV-1 envelope subtypes A and D on disease progression was investigated in 1045 Ugandan adults. In this case, subtype D was found to be associated with faster progression to death (P = 0.009) and with a lower CD4 cell count during follow-up (P = 0.001) compared with subtype A, after adjusting for CD4 cell count at enrolment. The conclusion was that, in Africa, envelope subtype D is associated with faster disease progression, compared with subtype A (Kaleebu *et al.*, 2002). Interestingly, this trend was observed previously (Alaeus *et al.*, 1999) in a study in which the authors did not show a significant difference between individuals infected with subtypes A-D, but there was a tendency for subtype D to progress faster than A: the mean CD4 lymphocyte count decline in D being twice that in A.

3.1.1.2 Vertical transmission

In addition to differences in disease progression, differences in transmission efficiency between subtypes have also been reported. An early study in Tanzania established the HIV-1 *gag* and *env* subtype and maternal risk factors of 51 matched transmitting and non-transmitting mothers. Transmission differences among genotypes revealed that mothers infected with HIV-1 subtypes A, C, or HIV-1 intersubtype recombinant viruses were more likely to transmit HIV-1 to their infants than mothers infected with HIV-1 subtype D. Lower CD4 cell counts at enrolment were associated with transmission, but CD4 cell counts within each genotype did not explain differences in transmission among HIV-1 genotypes. This study provided the first evidence that HIV-1 genetic subtypes may play a role in rates of vertical transmission in an African setting (Renjifo *et al.*, 2001). This was followed by a second study in Tanzania, in which transmission of viruses with different long terminal repeat (LTR) subtypes was investigated. Specifically, the LTR subtype was determined for 45 matched cases and controls; HIV-1 subtypes A, C, D and intersubtype recombinant sequences were identified. Statistical analysis showed that viruses containing subtype A or intersubtype recombinant LTRs were 3.2 and 4.8 times more likely to be transmitted from mother to infant than viruses with subtype D LTRs. Viruses containing subtype C LTRs were 6.1 times more likely

to be transmitted than those with subtype D LTRs. Differences in transmission were again shown to be independent of maternal CD4 at enrolment (Blackard *et al.*, 2001). More recently a study in Dar es Salaam, Tanzania, confirmed these observations by showing differences in timing for transmission from mothers to their infants. This much larger study of 253 HIV-1-infected infants was designed to examine the efficacy of vitamins in decreasing mother-to-child transmission. It was found, however, that there were significant differences in the distribution of transmission time according to subtype: a higher proportion of HIV-1 subtype C (*env*) was transmitted *in utero* than subtypes A or D (*env*) (Renjufi *et al.*, 2004).

In contrast to these data, however, a study performed in Kenya among 414 women for whom HIV-1 subtype and HIV transmission status were available, found that mother to child transmission rates were higher among women with subtype D (*env*/p24) compared with subtype A (Yang *et al.*, 2003). Furthermore, women with *env*/p24 subtype combinations D/D, D/A, and A/D had an increased risk of transmission than A/A women after adjusting for confounding factors. Not all studies, however, have reported subtype specific differences in vertical HIV-1 transmission. A study of 130 HIV positive mothers in Nairobi, Kenya, found no difference in transmission frequency between HIV-1 subtypes A and D (Murray *et al.*, 2000). In addition, a study of 31 HIV-1-seropositive pregnant Tanzanian women showed that, among eight transmitter mothers, four were infected with HIV-1 subtype A, one with HIV-1 subtype C, none with HIV-1 subtype D and three with HIV-1 subtype recombinant A/C. The authors concluded that these findings show no significant differences in the mother-to-child transmissibility of HIV-1 subtypes A, C and D and detected recombinants forms, although the study size in this case is particularly small (Tapia *et al.*, 2003). Perhaps the failure to detect subtype D transmission may in fact support the findings of the research by Renjufi *et al.*, in Tanzania.

Thus, several interesting hypotheses can be proposed regarding the relationship between viral subtypes and the perinatal transmission of HIV, in particular subtype A and D viruses. Inter-subtype genetic variation may bring about changes in receptor affinity, mediated by the *env* gene, and increased transcriptional activation, mediated by changes in the LTR and *tat*, accounting for these observations. More specifically, the higher rate of vertical transmission in mothers infected with subtype D compared to those infected

with subtype A could perhaps be explained on the basis of viral load, as the mean viral load among mothers infected with subtype D virus was often reported as being higher than that among mothers infected with subtype A (Yang *et al.*, 2003). The fact that this relationship between subtype and vertical transmission appears to be independent of viral load, however, suggests that there may be intrinsic properties of subtype D viruses that render them more infectious. Women who were either concordant for subtype D (D/D) or discordant with subtype A (D/A or A/D) had high transmission rates, indicating that it is not recombinant subtypes themselves that result in higher rates of transmission in this study population (Yang *et al.*, 2003), but rather the presence of the subtype D sequence in the genome. This suggests that viruses with subtype D are either more virulent or have a better fitness capacity or altered cellular tropism for placental cells in the Kenyan mothers studied. An alternative explanation could be that subtype A viruses are less virulent or less fit and may be transmitted less efficiently from mother to child. Most CRF in Africa have at least some subtype A structure (i.e. CRF01_AE, CRF02_AG) and if such recombination with subtype A occurs due to reduced efficiency of replication and transmission in subtype A, then it is likely that the HIV-1 subtype A epidemic may be replaced by one consisting of non-A subtypes or recombinant forms.

The estimated vertical transmission rate in Kenya of 25% (Datta *et al.*, 1994) is comparable with rates in other parts of the world: North American and European subtype B transmission rates are in the range of 19-25% (Matheson *et al.*, 1995); in South Africa subtype C virus perinatal transmission rate is 24% (Bobat *et al.*, 1997); and in Thailand subtype AE virus perinatal transmission rate is 24.2% (Shaffer *et al.*, 1999). Thus, given that vertical transmission rates from around the world are similar despite the prevalence of different subtypes this suggests that subtype may not, globally, have an effect on the rate of transmission of HIV-1 from mother to child. Furthermore, this may indicate that viral adaptation within the human population has resulted in equivalent transmission capacity, irrespective of viral subtype, due to host pressures. Differences in host ethnicity, environmental factors and viral genotype may all contribute to the subtle differences in transmission frequency recorded. It seems likely that HIV-1 subtype, viral load, host ethnicity, environmental factors and the likelihood of vertical transmission are intimately linked.

3.1.1.3 Horizontal (sexual) transmission

Many factors that impact on the sexual transmission of HIV have been described in epidemiological studies. These include: the distribution/frequency of risk behaviours; the prevalence of co-factors for HIV transmission (other STDs); immune factors; and non-immune factors such as male circumcision (Hu *et al.*, 1999). These factors, in combination with the challenge of quantifying incidence, have made attempts to study the relationship between likelihood of sexual transmission and HIV-1 sequence variation very difficult. Like the role of HIV sequence variation in vertical transmission of HIV-1, the importance of subtypes in sexual transmission is far from clear.

One of the best examples of a well-documented epidemic, however, was the independent introduction and subsequent spread of HIV-1 subtypes AE and B in Thailand. Whilst both subtypes were introduced at similar times it is subtype AE which has gone on to affect all risk groups within the population (Wasi *et al.*, 1995). Despite the many confounding issues in trying to relate genetic subtype to transmission characteristics, an observation as clear as this lead to the general hypothesis that certain subtypes of HIV-1, such as AE, may be transmitted more successfully through heterosexual intercourse. A cross-sectional study of HIV infected men and their partners, in Thailand, showed that there were higher rates of seroconcordance among couples infected with subtype AE compared to those infected with subtype B (Kunanusont *et al.*, 1995). Subsequent *in vitro* studies provided support for these findings as it was reported that subtype AE viruses from Thai heterosexuals grew more efficiently in vaginal epithelial Langerhans cells (LC), than subtype B viruses from U.S. homosexuals. This suggests that LC tropism is associated with the efficiency of heterosexual transmission of HIV (Soto-Ramirez *et al.*, 1996). The findings of this study, however, could not be repeated and further *in vitro* studies proved non-conclusive (Pope *et al.*, 1997, Dittmar *et al.*, 1997). The explanation as to how subtype AE predominates in Thailand thus remains highly controversial.

Furthermore, the rates of HIV-1 sexual transmission reported in different studies are quite variable. For example, a study in Thailand revealed a high per-act probability of heterosexual transmission, which was proposed to help explain the rapid spread of HIV-1 during the emergence of the epidemic in Thailand, and perhaps in other countries where HIV-1 transmission is predominately heterosexual (Mastro *et al.*, 1994).

However, as noted other studies have not reported similar transmission frequencies, irrespective of subtype, for either sexual (Mastro *et al.,* 1996) or vertical transmission (Shaffer *et al.,* 1999). Whilst some studies may support the notion that different HIV subtypes are differentially transmitted sexually, therefore, it is difficult to delineate what additional factors might contribute to the differences in the rates of transmission because of the variations in study methodologies and potential confounding factors.

3.1.1.4 Viral set point

The prognostic value of a single measurement of plasma virus load among patients infected with HIV-1 is widely recognized (Mellors *et al.,* 1997, Anastos *et al.,* 1999, Vlahov *et al.,* 1998, O'Brien *et al.,* 1996). Studies involving serial measurements of HIV RNA load, however, have provided data that argue both for and against a viral 'set point' theory, which proposes that virus load for an infected patient stabilises soon after HIV seroconversion, with the 'set point' achieved being a strong indicator of future disease progression (Katzenstein *et al.,* 1996; Bruisten *et al.,* 1997, Keet *et al.,* 1997; O'Brien *et al.,* 1998, Schacker *et al.,* 1998). Researchers have also hypothesised that an early dynamic between viral replication and the immune response may influence a person's longitudinal pattern of immunologic and virologic markers, as well as their clinical prognosis (Henrard *et al.,* 1995). It has also been reported that there may be differences in viral load at peak viraemia shortly after infection, before the set point is reached, between different HIV subtypes. There is therefore some debate as to whether different subtypes may be associated with certain 'set points', and thus certain phenotypes of clinical disease.

A study carried out by Rinke de Wit *et al.,* (2002) found that CD4 T-cell counts remained significantly lower and CD8 T-cell counts significantly higher in Ethiopian seroconverters compared with Dutch seroconverters. Viral loads were lower in Ethiopians in the first months after seroconversion, subsequently increasing to similar levels to the Dutch seroconverters. Furthermore, within the 20 Ethiopian seroconverters, 15 were infected with sub-cluster C' and 5 with sub-cluster C (Abebe *et al.,* 2000). Interestingly, viral loads were higher in sub-cluster C'-infected Ethiopians, suggesting that whilst subtype may have an influence on viral set point and therefore subsequent disease, there may be more subtle effects within genotypic variants of HIV which could be of biological importance.

A similar study looked at the difference between subtypes B and AE, again in Thailand (Hu *et al.*, 2001). 130 HIV-1-infected seroconverters (103 with HIV-1 subtype E and 27 with subtype B) were included in the study and serial HIV-1 RNA viral load, natural killer cell percentage, CD4 and CD8 lymphocyte counts were monitored over 24 months, post-seroconversion. Results showed that median viral RNA levels within 3 months of seroconversion were more than three times higher for persons infected with subtype AE than with subtype B, this difference decreasing over time such that viral loads were similar at 12, 18, and 24 months following seroconversion. The CD4 and CD8 lymphocyte counts were similar in infections with either subtype (Hu *et al.*, 2001).

To what extent these differences are influenced by the interaction between the virus genotype and the host environment is not known. Hu *et al.* hypothesise that higher viral loads associated with subtype AE may result from inter-subtype biological differences. Characterisation of subtype B and AE viruses from this study cohort showed major inter-subtype differences in the proportions of different envelope V3 motifs as well as predicted co-receptor usage and phenotype from genetic sequence data (Subbarao *et al.*, 2000). Furthermore, reported inter-subtype differences in the function of long terminal repeat (LTR) transcriptional promoters suggest that subtype AE may replicate more efficiently than subtype B in certain circumstances (Jeeninga *et al.*, 2000). The epidemiological dynamics of transmission in Bangkok, however, were also proposed to have contributed to this phenomenon. That is, subtype AE was introduced into Thailand after subtype B, at a time when there was a higher incidence of intravenous drug users in the affected population. Accordingly, one might predict a higher proportion of persons with subtype AE to have been more recently infected and hence have higher mean viraemia (characteristic of acute disease). Thus, if these recently infected persons in the study described were serving as source partners, then the subtype AE reservoir of infection would be richer in higher viral loads than the subtype B reservoir (Hu *et al.*, 2001). As higher viral load has been associated with higher rates of transmission through a variety of modes (Mastro *et al.*, 1998, Shaffer *et al.*, 1999) if a person is infected by someone with subtype AE they would be more likely to have encountered someone with a higher viral load (thus inoculum), which in turn would result in higher initial viral loads in the seroconverter, thus perpetuating the spread of HIV-1 subtype AE.

3.1.1.5 HIV-1 and HIV-2

When considering the importance of HIV sequence variation in terms of disease progression and transmission (both sexual and vertical), the difference between HIV-1 and HIV-2 provides a very useful comparison. That is, between these two viruses there is approximately 50% sequence divergence and whilst HIV-2 shares the same modes of transmission as HIV-1 (DeCock *et al.*, 1993, Poulsen *et al.*, 1989, Gnaore *et al.*, 1989), the distribution of HIV-2 infection mainly remains restricted to west Africa where prevalence rates have been stable over time, in contrast to the global distribution and prevalence of HIV-1 (Clavel *et al.*, 1986a,b). Further to these observations, numerous studies have demonstrated that, unlike between the subtypes of HIV-1, there are clear and appreciable differences in transmissibility between the two HIV types, with much lower rates of sexual and mother-to-infant transmission for HIV-2 than for HIV-1: whereas approximately one-quarter of infants born to HIV-1-infected women around the world become infected in the absence of preventive interventions (Oxtoby, 1990, European Collaborative Study, 1991, Halsey *et al.*, 1990) approximately 1% of HIV-2-infected mothers transmit the infection to their newborns (Andreasson *et al.*, 1993, The HIV Infection in Newborns French Collaborative Study Group, 1994, Adjorlolo-Johnson *et al.*, 1994). It may be that, if the relative differences in transmissibility were smaller, these studies might not have been able to demonstrate significant differences. Natural history studies have also demonstrated that HIV-2 has a reduced virulence compared to HIV-1 (Marlink *et al.*, 1994), and HIV-2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2 (Popper *et al.*, 1999).

Accordingly, it may be that using epidemiological studies to identify the differences between HIV subtypes (which show approximately 25-30% sequence divergence) may be too insensitive to distinguish sequence variation effects from the myriad of other factors that modulate HIV-1 transmission and disease progression, at the population level, unless extremely large and well-controlled studies are undertaken. If using sequence generated for resistance testing, such studies would be compounded by the methods for subtype assignment of *pol* sequence currently available which, if not incorporating more complex phylogenetic approaches, must be questioned as to their accuracy (Chapter 2).

3.1.2 *In vitro* studies

Given the inconclusive link between HIV-1 genetic variation and disease phenotype, the simplest hypothesis may be investigated: that genetically variable strains of HIV will grow differently in cell culture. If this is true, and the findings agree with previously published observations regarding *in vivo* differences between HIV subtypes, this could lend weight to the argument that subtype is an important causal factor in the phenotype of HIV disease. Accordingly, several *in vitro* studies have attempted to characterise the phenotypic manifestations of genotypic variation. Studies looking at the relationship between HIV-1 subtype, coreceptor usage and *in vitro* viral pathogenicity, and HIV subtypes B and AE *in vitro* transmissibility have been described. Study of the HIV-1 LTRs, *PR, env* and accessory genes of different subtypes, however, have also generated interesting data.

3.1.2.1 The HIV-1 LTR

The LTR is a major determinant for virulence in several animal retroviruses and minor changes or rearrangements within the transcription factor binding sites (TFBS) have been shown to have a significant impact on cell tropism and pathogenicity (Carvalho *et al.*, 1993, Maury *et al.*, 1998). In general the effect of HIV-1 LTR subtype on viral infectivity, virulence and fitness, however, is not known. Small differences in LTR transcriptional activity between subtypes have been demonstrated (Jeeninga *et al.*, 2000, Montano *et al.*, 1998, Naghavi *et al.*, 2001, Quivy *et al.*, 2002). Whether these differences in LTR activity result in subtype-specific differences in replication rate, however, was only addressed recently in a study that aimed to determine to what extent LTR subtype and the cellular environment contributes to viral replication (van Opijnen *et al.*, 2004). Specifically, the importance of the LTR on interactions between six cellular environments and nine viral genotypes (A-G and AG) with an isogenic set of HIV-1 molecular clones containing a subtype-specific LTR was determined. By conducting pair-wise competition experiments between all virus variants in all cellular environments it was shown that there are significant differences in replication rate between the subtypes and that the LTR-determined viral fitness depends on both the host cell type and the activation state. Subtype AE for example, had the strongest competitive ability and out-competed all other viruses, which ranked as sub-subtype C2 being the second best, followed by subtypes A, C1, G, D, AG, B, and F. Conversely,

when competition experiments were performed in a SupT1-tumour necrosis factor (TNF)-environment, subtype AE changed from being the strongest competitor with the highest replication rate in SupT1 cells to being the worst competitor with the lowest replication rate. Thus, these data suggest that a given mutation can have a positive fitness effect in one cellular environment and a negative fitness effect in another (van Opijnen *et al.*, 2004). One of the consequences of this study, therefore, is that it may not be possible to assign a single fitness value to a particular subtype. The results clearly demonstrate a single replication or competition experiment may not provide an accurate estimate of fitness, because fitness (as determined by the LTR promoter) depends on the interplay between the genetic structure of the virus and the cellular environment. It is therefore not possible to tell which subtype is better, but only which subtype is better in a particular environment.

3.1.2.2 Protease and Gag

A study of HIV-1 subtype A and C protease enzymes in the presence of several protease inhibitors revealed that the biochemical fitness of subtype A and C proteases is higher than that of the B subtype protease (Velazquez-Campoy *et al.*, 2001). Considering the fact that currently available inhibitors were designed against the B subtype protease, this finding is perhaps not surprising. It is difficult, however, to translate the difference in biochemical fitness observed into a meaningful biological significance.

Investigations regarding differences between HIV-1 subtypes have also focussed on the protease cleavage sites (CS). Specifically, the prevalence and patterns of polymorphisms in the Gag, Gag-Pol, and Nef CS of subtype C compared to those in non-C HIV subtypes, was assessed. Significant inter- and intrasubtype differences in CS, especially in the p2/NC, transframe protein/p6pol, and p6pol/PR sites, were identified (de Oliveira *et al.*, 2003): 58.3% of the 12 CS were significantly more diverse in C than in B viruses. The authors propose that natural variation at subtype C CS may play an important role, not only in regulation of the viral cycle but also in disease progression and response to therapy.

3.1.2.3 Env

The reason that many *in vitro* studies have focussed on subtype C is that the outgrowth of subtype C in southern Africa, India, and China has suggested that subtype C isolates

may be more fit *in vivo*. In a recent study by Ball *et al.*, nine subtype B and six subtype C HIV-1 isolates were added to peripheral blood mononuclear cell cultures for a complete pair-wise competition experiment. It was found that all subtype C HIV-1 isolates were less fit than subtype B isolates ($P < 0.0001$), and that intrasubtype variations in HIV-1 fitness were not significant. Increased fitness of subtype B over subtype C was also observed in primary CD4+ T cells and macrophages, but not in skin-derived human LC, and analysis of the retroviral life cycle during several B and C virus competitions indicated that the efficiency of host cell entry may have the greatest impact on relative fitness. Furthermore, for a recombined subtype B/C HIV-1 isolate, higher fitness mapped to the subtype B *env* gene rather than the subtype C *gag* and *pol* genes. This suggests that subtype B and C HIV-1 may be transmitted with equal efficiency (via LC) but that subtype C isolates may be less fit following initial infection (as shown in T-cell and macrophage experiments) and thus may lead to slower disease progression (Ball *et al.*, 2003).

For HIV-1 and HIV-2, however, clear differences in the effect of gp120 and gp105 (HIV-2 envelope glycoprotein) have been proposed to contribute to the distinct pathogenicity of the two viruses. Specifically, studies have shown that gp105 is more effective in inducing β-chemokines than gp120 (Neoh *et al.*, 1997), and it has been suggested that HIV-2 gp105 is able to use a wider range of co-receptors than HIV-1 gp 120 (Bron *et al.*, 1997). Perhaps more importantly, however, is the observation that gp105 is more inhibitory to T cell proliferation and the up regulation of CD40 ligand and OX40, than gp120, in the absence of significant induction of apoptosis. This marked immunosuppressive property of the HIV-2 envelope could be beneficial to the host by interfering with the heightened state of cellular activation that characterises HIV-1 infection and by limiting the bursts of viral replication (Cavaleiro *et al.*, 2000).

3.1.2.4 Accessory genes

Rev and Vpu

In addition to viral studies which have shown that subtype C has distinct genetic and phenotypic properties that differentiate it from other HIV-1 subtypes, the presence of a prematurely truncated Rev protein and a 5-amino-acid insertion in Vpu have been identified (Gao *et al.*, 1998). Specifically, all subtype C viruses contain a stop codon in the second exon of Rev, which would be predicted to shorten this protein by 16 amino

acids, and a 15-bp insertion at the 5' end of the *vpu* gene extends the putative membrane-spanning domain of the Vpu protein by 5 amino acids. Although these changes are unlikely to alter the function of the respective gene products in a major way it is possible that they could influence their mechanism of action in a subtle, but biologically important way. Furthermore, a study of Vpu sequences from 101 strains of HIV-1 of various subtypes revealed that all subtype C sequences had a conserved LRLL motif at the C terminus that was also found in A/C intersubtype recombinants (Mccormick-Davis *et al.*, 2000). In addition to serving as a useful epidemiological marker to identify subtype C isolates, this motif may be of interest as dileucine motifs have been demonstrated to mediate lysosomal targeting and endocytosis of CD3 chains, IgG Fc receptors, and mannose 6-phosphate receptors, and to be involved in Nef mediated CD4 down regulation (Hunziker *et al.*, 1996, Craig *et al.*, 1998). Subtype C viruses, therefore, may be distinct in their capacity to modulate T-cell receptor activity (Palacios and Weiss, 2004), impair cell-mediated immunity (Kedzierska *et al.*, 2002) and traffic acid hydrolases to endosomal compartments (Ghosh, Dahms and Kornfeld, 2003), respectively, which may in turn influence the capacity of this viral subtype to persist and replicate within a host.

Tat

A study has also revealed that subtype C Tat exhibits higher transcriptional activity from the HIV-1 long-terminal repeat (LTR) in a human T-cell line, compared to subtypes B and AE (Kurosu *et al.*, 2002). This higher activity of subtype C Tat was associated with two amino acid changes compared to subtypes B and AE, within and close to the basic domain. As Tat nuclear localization activity was not affected, these data suggests that there may be a significant advantage for the high Tat activity on subtype C replication.

3.1.3 Study of HIV-1 primary isolates *in vitro*

As outlined in the introduction to this thesis, when HIV subtype is thought of as a 'risk' factor, many epidemiological studies are flawed in that even if subtype is a true causal factor in disease phenotype, it may be indistinguishable from other factors that contribute to a particular outcome. As a result, the work covered in this chapter represents a conceptually more simplistic approach to the study of different HIV-1

subtypes, *in vitro,* with the goal of characterising basic differences between different viral genetic variants. The cornerstone to this work, therefore, was the production of a panel of HIV-1 group M, O and HIV-2 primary isolates. Following creation of an extensive viral resource for *in vitro* studies, experiments were performed to characterise the effect of viral genetic variation, on *in vitro* infection phenotype.

3.1.3.1 Primary isolate viruses

The viruses chosen for study are summarised in Table 3.0. All were low passage primary isolates, as the aim was to investigate the differences between the growth of different subtype viruses *in vitro,* using material as biologically relevant as possible. This was important in order to avoid the confounding issues of studying heavily tissue culture or cell-line adapted viral strains. To date few studies have directly compared a large number of un-modified primary isolates, in the same experimental system. Virus stocks were generated in PBMCs and the viruses used for experiments were only two passages away from the original sample. Whilst ideally viruses of different subtype would have been directly isolated from patients prior to this study, given the impracticality of this and the fact that for several of the viruses chosen sequence and phenotypic information was already available, this approach was considered to be optimal.

HIV-1 group M subtypes A, AE, B, C, D, and F were selected as they represent the diversity of viral genetic variants which play a significant role in the HIV-1 pandemic. A group O virus was also chosen, as there was little information available regarding the phenotypic difference between Group M and O viruses, either *in vitro* or *in vivo,* despite considerable genotypic divergence. Conversely, HIV-2 ETP was included in this study as the differences between HIV-1 and HIV-2, both *in vitro* and *in vivo* have been well documented.

3.1.3.2 Methods of study

Environment-specific effects on growth phenotype have been clearly demonstrated for the different HIV-1 group M subtypes (van Opijnen *et al.,* 2004). As the aim of this chapter was to better characterise the influence of HIV genotypic variation on growth phenotype *in vitro,* a system in which the growth of diverse viruses could be experimentally compared was required. Ideally, the cells used must be infectable with

both X4- and R5-tropic viruses, and contain a form of reporter system that will be representative of infection, over time, but would not be subject to bias due to infection with genetically divergent HIV-1 subtypes, and HIV-2. It was therefore the aim of this chapter to firstly produce and sequence verify a panel of low passage HIV primary isolates (Table 3.0). Secondly, their growth characteristics *in vitro* were assessed in PBMCs and on a T-cell line, and finally a reporter cell line was developed to further characterise these isolates. Its utility in characterising the ART sensitivity of HIV primary isolates was also demonstrated.

3.1.4 Aims

It is a widely held belief that it is unlikely that a single characteristic such as viral subtype can account for significant differences in HIV disease progression (Hu *et al.*, 1999). The association of subtype with certain viral phenotypic traits, however, may bear further exploration, as in some cases subtype may be a surrogate marker for a particular viral phenotype of importance in terms of disease phenotype. With this is mind it was the aim of this chapter to characterise phenotypic determinants of different HIV-1 subtypes and HIV-2 *in vitro,* in a number of different cellular environments, bearing in mind the significance of any such differences *in vivo.*

Table 3.0 Summary of HIV-1 and -2 primary isolates

Subtype	Ref.	Source	CoRe	SI/NSI	Reference
A	92UG037	Entebbe, Uganda, 1992. Asymptomatic seropositive 31-year-old female.	ND	ND	The WHO Network for HIV Isolation and Characterisation, 1994
AE	92TH001	Thailand, 1992. Seropositive individual.	ND	ND	The WHO Network for HIV Isolation and Characterisation, 1994
B	SF162	USA, 1989. Isolated from CSF of AIDS patient.	CCR5	NSI	Cheng-Mayer *et al.*, 1989
C	92BR025	Brazil, 1992. Seropositive individual.	ND	ND	The WHO Network for HIV Isolation and Characterisation, 1994
D	92UG001	Uganda, 1992. Seropositive individual.	ND	SI	The WHO Network for HIV Isolation and Characterisation, 1994
F	93BR020	Brazil, 1993. Seropositive individual.	ND	ND	The WHO Network for HIV Isolation and Characterisation, 1994
O	BCF06	France, 1994. PBMCs of 22yr old AIDS patient.	CXCR4	SI	Simon *et al.*, 1998
HIV-2	ETP	Portugal, 1999. AIDS patient.	dual	SI	Reeves *et al.*, 1999

Ref: reference that identifies each isolate, assigned at source
CoRe: coreceptor usage of the isolate in question - ND indicates Not Determined
SI/NSI: refers to the cytopathic characteristics of each virus - SI indicates Syncytium Inducing, NSI indicates Non Syncytium Inducing

3.2 Methods

3.2.1 Cells and viruses

Peripheral blood mononuclear cells (PBMCs) from HIV-seronegative blood donors were obtained by Ficoll-Hypaque density gradient centrifugation of buffy coats from two separate donors (using Lymphoprep, Nycomed, Norway). Prior to HIV-1 infection, cells from both donors were mixed and stimulated with 0.5 μg phytohaemagglutinin/ml (PHA, Gibco BRL, UK) for 3 to 4 days and maintained in Roswell Park Memorial Institute (RPMI) 1640-2 mM L-glutamine medium, supplemented with 10% foetal bovine serum (FBS, BioWest, France), 10 units (U) of interleukin-2/ml (IL-2), 100 U/ml penicillin and 100 μg/ml streptomycin (P/S, Invitrogen, UK). SupT1.R5 (ST1-R5) and CEM.R5 cell lines were maintained in RPMI 1640-2 mM L-glutamine medium, supplemented with 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin and 1 μg of puromycin/ml. NP2/CD4/CXCR4, NP2/CD4/CCR5 and HEK 293-T cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM, Invitrogen, UK) supplemented with 5 or 10% FBS, 100 U/ml penicillin and 100 μg/ml streptomycin, NP2 cells with an additional 1 μg of puromycin/ml. CVG cells - clone 37 (CVG-37) were propagated and maintained in RPMI 1640-2 mM L-glutamine medium, supplemented with 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin, 1 mg G418/ml (Invitrogen, UK) and 1 μg puromycin per ml (Invitrogen, UK). Parental CEM-G cells were cultured in the presence of G418 alone. All cells and cell lines were incubated at 37°C in 5% CO_2. References are given in Table 3.1.

Table 3.1 Cell lines: Description of cell lines used in this study.

Cell line	Description	Reference
HEK 293-T	293 cells containing the temperature sensitive gene encoding SV40 T-antigen	DuBridge *et al.*, 1987
NP2.CXCR4	CD4 and CXCR4 transduced human glioma cell line	Soda *et al.*, 1999
NP2.CCR5	CD4 and CCR5 transduced human glioma cell line	Soda *et al.*, 1999
CEM.R5	Human T-lymphoblastoid cell line transduced with CCR5	Trkola *et al.*, 1999
ST1-R5	Non-Hodgkin's T-cell lymphoma cell line transduced with CCR5	Means *et al.*, 2001
CEM-G	Human T-lymphoblastoid cell line. Contains plasmid encoding LTR-driven humanised S65T GFP	Gervaix *et al.*, 1997
CVG-37	Human T-lymphoblastoid cell line. Contains plasmid encoding LTR-driven humanised S65T GFP, and stably expresses CCR5	-

3.2.2 Production of primary isolate stocks

Six primary HIV-1 isolates (Table 3.0) were obtained from the National Institute for Biological Standards and Control (NIBSC, Potters Bar, UK): subtypes A (92UG037); AE (92TH001); C (92BR025), D (92UG001), F (93BR030) and O (BCF06). Subtype B isolates (SF162) and HIV-2 isolate (ETP) were kind gifts from the Wohl Virion Centre (UCL, London). To prepare stocks, 1 ml of each virus was added to $8x10^6$ PHA- and IL-2-treated PBMCs. After 2 hours (h) of incubation at 37°C, cells were resuspended in complete medium (10^6 cells per ml). Approximately 3-5 days post infection (virus dependent), infected cells were co-cultivated with 5x the original number of PBMCs (2 h, 37°C) prior to resuspension in complete medium (10^6 cells per ml). Viruses were harvested as appropriately by centrifugation at 1200 rpm/5 minutes, removal of virus containing supernatant, recentrifugation to clarify and filtration though a 0.45 μM filter (Sartorius, UK). Virus was aliquoted into cryovials (Nunc, USA), snap frozen and stored in vapour phase liquid nitrogen until use.

3.2.3 DNA purification

DNA was purified from PBMCs infected with all HIV-1 primary isolates using a Qiagen QIAamp DNA mini kit (Qiagen, UK). $2x10^6$ infected PBMCs cells were collected after virus harvest resuspended in 200 μl PBS with 20 μl proteinase K (Qiagen, UK). Following lysis, samples were incubated at 56°C for 30 minutes, then applied to a QIAamp spin column and centrifuged at 14,000g for 1 minute, washed twice with appropriate buffers and eluted in a final volume of 100 μl. The concentration of the purified DNA was calculated using a NanoDrop spectrophotometer (NanoDrop Technologies, Rockland DE, USA).

3.2.4 Polymerase chain reaction (PCR) amplification

PCR was performed to amplify sections of the group specific antigen (Gag), protease (PR) and reverse transcriptase (RT) genes from DNA extracted from HIV-1 infected PBMCs (Dyad thermal cycler, MJ Research Inc. Reno NV, USA). A pair of degenerate primers were designed to amplify these targets from HIV-1 group M subtypes and a separate set to amplify each from HIV-1 group O (Table 3.2). The constituents of each reaction (A) and reaction conditions (B) were:

(A) Reagent	Volume
DNA (approx. 50 ng)	2 µl
Primers (10pMol/µl)	2 µl each
PCR buffer (10x)	5 µl
$MgCl_2$ (25 mM)	4 µl
dNTPs (10 mM)	1 µl
Taq DNA polymerase (5 U/µl)	0.25 µl
Distilled water	35.75 µl

(Promega, UK).

(B) Temperature	Time
94°C	120 seconds
94°C	45 seconds
55°C	45 seconds
72°C	60 seconds
Steps 2-4 an additional 40 times	
72°C	120 seconds

3.2.5 PCR purification

10 µl of the PCR reaction was separated on a 1% w/v agarose gel in 1x Tris-borate-ethylenediaminetetraacetic acid (EDTA) (TBE) containing 0.2 µg/ml ethidium bromide (Sigma, UK) at 80V. The band of the correct size was excised from the gel and DNA purified according to the QIAquick gel extraction protocol (Qiagen, UK). In short, following solubilisation of the gel slice at 50°C for 10 minutes, isopropanol was added, the solution vortexed and applied to a QIAquick column by centrifugation (14,000g, 1 minute). After washing with solubilisation and wash buffers the DNA was eluted in 30 µl of water.

Table 3.2 PCR primers used in this study.

Primer	Sequence (5' - 3')	Target
T7	TAATACGACTCACTATAGGG	pGEM T-easy
SP6	TATTTAGGTGACACTATAG	pGEM T-easy
Env F-I	CAACTCAACTRCTGTTRAATGG	HIV-1 env
Env R-I	GTAATTTCTRGGTCCCCTCCTGA	(inner primer)
Env FO-I	GTACTCAACTAATATTAAATGG	HIV-1 env gp.O
Env RO-I	GTTACCTCTGGATCTCCACCACT	(inner primer)
Env F	GYAAGAATGTCAGCACAGTACAATGYACACATGG	HIV-1 env
Env R	GWATTRCARTAGAAAAATTCYCCTCCACAATT	(outer primer)
Env F-O	GCAAAAAYATTACAGTAGTTACTTGTACACATGG	HIV-1 env gp.O
Env R-O	GTGTTACAATARAAGAATTCTCCATGACAGTT	(outer primer)
Gag F	GGTGCCAGAGCGTCARTATTAAGHGGGGGAGAATTAG	HIV-1 gag
Gag R	GTAGTTCCTGCTATRTCACTTCCCCTTGGTTCTC	
Gag F-O	GGTGCCAGTGCGTCTGTGTTGACAGGGAGCAAATTGG	HIV-1 gag gp.O
Gag R-O	GTTGTYCCAGCAATGTCACTTCCTGTTGGTTCCC	
PR F	CAGARCAGACCRGAGCCAACAGCCCCRCC	HIV-1 PR
PR R	GGKCCATCCATTCCTGGCTTYAATKTTACTGG	
PR F-O	CAGAGACAAGTGTCCCCATCAGCCCCACC	HIV-1 PR gp.O
PR R-O	GGTCCATCCATTCCTGGTTTTAGTTTTACTGG	
RT F	CAACATAATTGGAMGAAATMTGTTGACTCAGMTTGG	HIV-1 RT
RT R	CTRTGYTGCCTTATTTCTAARTCAGAYCCTACATAC	
RT F-O	TAATATTATTGGAAGAAACATATTGACAGGATTAGG	HIV-1 RT gp.O
RT R-O	CTATGTTCTGTCAGGGGTAAATCTGATCCTACATAT	
SS-fwd	GGCTAACTAGGGAACCCACTG	HIV-1 strong stop
SS-rev	CTGCTAGAGATTTTCCACACTGAC	DNA
SSO-fwd	GGCCTCTAGCTGAACCCGCTG	HIV-1 gp.O strong
SSO-rev	CTGCTAGAGATTTTCTGCTTCAGT	stop DNA

3.2.6 Preparation of competent bacteria

XL-1 Blue *Escherichia coli* (Stratagene, UK) were streaked onto Luria-Bertani (LB)-agar plates containing tetracycline (10 μg/ml) and incubated at 37°C for 16 h. A colony was picked and used to inoculate 3 ml LB-broth (with tetracycline) and shaken at 37°C for 16 h. This was added to 500 ml LB-broth and shaken at 37°C until the absorbance at 600 nm was 0.6 (around 3 h). The culture was then put on ice to cool for 10 minutes. The bacteria were pelleted at 3000 xg for 10 minutes at 4°C and resuspended in 15 ml of 100 mM calcium chloride (4°C). After incubation on ice for 30 minutes, the bacteria were centrifuged again and then resuspended in 2.5 ml of 100 mM calcium chloride containing 15% glycerol by volume (4°C). The bacterial suspension was frozen at -80°C in 200 and 400 μl aliquots.

3.2.7 Cloning

PCR products were cloned into pGEM-T Easy (Promega, UK). 1 μl of DNA was ligated into 50 ng of pGEM-T Easy using 3 U of T4 DNA ligase and 5 μl rapid ligation buffer (2x) in 10 μl final volume at room temperature (RT) for 2 h or 4°C overnight. The ligation reaction was cooled to 4°C and added to 200 μl of competent XL-1 Blue *E.coli*. After incubation at 4°C for 30 minutes, the bacteria were subjected to heat-shock at 42°C for 45 seconds and then placed immediately on ice. 400 μl LB-broth was added and the bacteria incubated at 37°C for 30 minutes. The bacterial suspension was plated onto LB-agar containing 100 μg/ml ampicillin, 0.5 mM isopropylthio-β-D-galactoside (IPTG) (Sigma, UK) and 80 ng/ml 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) (Promega, UK). The plates were incubated at 37°C for 16 h.

3.2.7.1 Screening for positive colonies

Bacterial colonies transformed with vector containing PCR product insert were identified by their white colour due to disruption of the LacZ open reading frame. Colonies were screened by PCR for the presence of plasmid inserts of the correct size. The ingredients of the PCR were as in section 3.2.4 except in half-volumes. The primers used were T7 and SP6 (Table 3.2). The conditions used were:

Temperature	Time
94°C	5 minutes
94°C	30 seconds
55°C	30 seconds
72°C	60 seconds
Steps 2-4 an additional 29 times	
72°C	5 minutes

Positive colonies were added to 3 ml LB-broth plus amplicillin (50 μg/ml) and grown at 37°C for 16 h.

3.2.7.2 Purifying plasmid DNA

500 μl of bacterial suspension was frozen at –80°C in sterile 20% glycerol by volume. Plasmid DNA was purified from the remaining bacteria using QIAprep miniprep protocol (Qiagen, UK), a modified alkaline lysis method (Birnboim and Doly, 1979). 1.5 ml of culture was pelleted at 14,000xg for 10 minutes and the bacteria resuspended, lysed for no more than 5 minutes, the lysis then halted by neutralisation. Cell debris was pelleted by centrifugation at 14,000 xg for 10 minutes and the DNA-containing supernatant applied to a QIAprep column. The solution was passed through the column by centrifugation (14,000 xg, 1 minute), washed as directed and plasmid DNA eluted with 50 μl distilled water (average final concentration around 300 ng/μl).

3.2.7.3 DNA sequencing

The identities of clones were also verified by DNA sequencing, using an automated application of the chain-termination method (Sanger *et al.*, 1977). 1.5 μl of plasmid DNA (or approximately 500 ng) was added to 4 μl of Beckman QuickStart mix (containing dATP, dCTP, dTTP, dITP, ddUTP, ddGTP, ddCTP, ddATP and thermosequenase DNA polymerase in Tris buffer, pH 8.9) and 5 pmoles of T7 or Sp6 primer. The solution was placed in a thermal cycler and subjected to the following conditions:

Temperature	Time
94°C	3 minutes
96°C	20 seconds
50°C	20 seconds
60°C	4 minutes
Steps 2-4 an additional 29 times	

DNA was precipitated from the sequencing reaction by adding 2 µl sodium acetate (1.5 M), 2 µl EDTA (50 mM, Sigma, UK) and 1 µl glycogen (20 mg/ml, Beckman Coulter, UK) in a final volume of 20 µl, vortexing, and then adding 60 µl 95% (v/v) ethanol/water (-20°C) and leaving on ice for 10 minutes. DNA was pelleted by centrifugation at 14,000g for 15 minutes at 4°C. The supernatant was removed, the pellet washed with 200 µl 70% ethanol/water (-20°C), and dried at RT. DNA was resuspended in 40 µl de-ionised formamide (JT Baker, USA) for 15 minutes and transferred to a 96-well plate. The sequences were determined using an automated capillary DNA sequencer (Beckman Coulter, UK) as per the manufacturer's instructions. Sequence data were analysed using BioEdit software v.5.0.9 (Hall, 1999).

3.2.8 Immunostaining of HIV-2- and HIV-1-infected cells and calculation of titre

The immunostaining method has been described previously (Clapham *et al.,* 1992). Briefly, serial dilutions of virus were made and added to confluent NP2/CD4/CXCR4 or NP2/CD4/CCR5, depending on viral coreceptor usage. These were incubated at 37°C, 5% CO_2 for 3 h, virus then removed by washing with PBS and replaced with medium. After three days methanol-acetone (1:1)-fixed cells infected with HIV-2 or HIV-1 were immunostained with HIV-2 human serum diluted 1/4,000 or anti-HIV-1 p24 monoclonal antibody diluted 1/40 (1:1 mix of EVA 365 and 366 from the Medical Research Council AIDS Reagent Program, Potters Bar, United Kingdom). Second-layer ß-galactosidase conjugates of goat anti-human immunoglobulin G (HIV-2) or goat anti-mouse immunoglobulin G (HIV-1) were used to detect first-layer antibodies at a dilution of 1:400 (Southern Biotechnology Associates, Inc., Birmingham, Ala.). Infected cells were stained blue with X-Gal in PBS containing 3 mM potassium ferricyanide, 3 mM potassium ferrocyanide, and 1 mM magnesium chloride. Foci of infection, which stained blue, were counted using light microscopy and virus infectivity was estimated as FFU per millilitre.

3.2.9 Production of molecular clone-derived virus, NL4-3 and Yu2

293T cells were passaged 48 h prior to transfection. In a 10-cm^2 tissue culture dish, a semiconfluent (40 to 60%) monolayer was seeded 24 h later. The transfection reagent FuGENE6 was used according to the manufacturer's instructions. Briefly, a total of 4 µg of DNA was combined with 200 µl of serum-free Optimem (Gibco, Invitrogen). 15 µl

of FuGENE6 was added directly and the suspension incubated for 15 minutes at RT. The volume was increased to 1 ml with complete DMEM. The transfection mixture was added to the cells and incubated at 37°C, 5% CO_2 in a humidified incubator. Supernatants were harvested after 48 and 72 h.

3.2.10 Preparation and titration of CCR5 lentivirus

CCR5 cDNA was previously subcloned into pBABE-puro. Amphotropic virus stocks were prepared by cotransfecting 293T cells with the pBABE-puro.CCR5 construct, a VSV-G envelope expression vector (pMDG), and gag-pol expression vector (CMV-I) (all kindly donated by Dr. G Towers, UCL). Briefly, 293T cells were passaged 24 h prior to transfection in a 10-cm^2 tissue culture dish to 25% confluence. Vectors were prepared in 15 µl Tris-EDTA and added to 200 µl serum-free Optimen, containing 18 µl Fugene6, and incubated for 15 minutes at RT. Volume was increased to 1 ml with complete DMEM. The transfection mixture was added to the cells and incubated at 37°C, 5% CO_2 in a humidified incubator. Supernatants were harvested after 48, 72 and 96 h and titre calculated by TE671 titrations. Briefly, TE671 cells were cultured overnight in 6 well plates (Costar) at a density of 1 x 10^5 cells/well. The following day cells were exposed to 1 ml of serial dilutions of CCR5 lentivirus (10^{-1} to 10^{-6}) in complete DMEM supplemented with 5 µg polybrene per ml, and incubated at 37°C/5% CO_2 for 48 h. Virus was then removed and replaced with complete DMEM supplemented with 1µg puromycin per ml. After an additional 48 h incubation at 37°C/5% CO_2, media was replaced and 48 h later colonies could be visualised, counted and titre calculated.

3.2.11 Establishment and characterisation of CVG cells

CEM-G cells were added to one well of a 24 well plate (Costar) at a density of 1 x 10^5 cells/well. CCR5 lentivirus was used to infect these cells, by addition of 1:1 virus to complete RPMI (1 ml), incubation at 37°C/5% CO_2 for 4-6 h, and removal and replacement of 800 µl virus containing supernatant with complete RPMI. This was repeated each day for three days, the last day medium being supplemented with 1µg puromycin per ml. Outgrowth of resistant cells took approximately one week, which were then transferred to T75 flask (Nunc) and grown in complete puro-media for another week. Staining of the mixed population with a R-phycoerythrin (PE)-

conjugated mouse anti-human CCR5 antibody (BD Pharmingen) indicated very low expression of CCR5 (FACSCan flow cytometer (Becton Dickinson, Mountain View, CA), so limiting dilution plates were made in order to obtain clonal lines. These were maintained at 37°C/5% CO_2 for four weeks. 20 clones were selected (>2wks), 10 clones (>3wks) and an additional 10 clones (>4wks). FACS analysis was performed, and from these 13 were retained. These were grown up and analysed for CXCR4 (PE-conjugated mouse anti-human CXCR4 (Fusin), BD Pharmingen), CCR5 and CD4 (PE-conjugated mouse anti-human CD4, BD Pharmingen) expression. Isotype control mouse anti-human IgG_1-PE (Sigma, UK) was used. Two lines were selected that stably express human CCR5 together with CXCR4 and CD4, and have a low level of constitutive GFP expression.

As the clonal lines were previously stably transformed with GFP, which is under the control of HIV-1 LTR and is transactivated by HIV or SIV Tat upon infection with either X4 or R5 tropic HIV-1 or HIV-2, the cells should become green when infected. This was assessed by infection with NL4-3 (X4-tropic HIV-1) and Yu2 (R5 tropic HIV-1), at an MOI of 0.25. Briefly, molecular clone viruses NL4-3 and Yu2 were added to three clones of interest by incubation of 10^5 cells with virus at MOI 0.25, for 2 h at 37°C. Cells were then washed three times with PBS and resuspended in 1 ml complete medium. On days 2, 3, 4, 5, 6, 7 and 8 post-infection 200 µl of cells was removed from each infection and cell-free supernatants stored at –80°C prior to detection of p24. Cell pellets were resuspended in 500 µl 2% formolsaline and stored at 4°C in the dark, prior to FACs analysis. Photographs were also taken over this time period (using UV microscopy) to record visually detectable changes in GFP expression.

3.2.12 HIV-1 subtype timecourses

3.2.12.1 Replication assay

Six primary HIV-1 isolates (not 92TH001) and ETP were used for infection of PBMCs, ST1-R5 and CVG-37 cells. For ST1-R5 and PBMC infections, viruses were added at a multiplicity of infection of 0.01 ffu/cell and incubated at 37°C for 4 h in the presence of 5 µg per ml polybrene. For CVG-37 infections, viruses were added at a multiplicity of infection of 0.01 ffu/cell and spinoculated at 13000rpm/25°C for one hour, then

3000rpm/25°C for 1hour, followed by incubation in the presence of 5 μg per ml polybrene at 37°C for 2 h. In all cases cells were then washed three times with PBS and resuspended in the appropriate complete medium and six aliquots of 200 μl added to separate wells of a 96 well plate (1 x 10^4 cells/well). Over a period of several days post infection (up to two weeks) one well of cells infected with each of the eight viruses was removed and cell-free supernatants stored at –80°C prior to detection of p24. Cell pellets were resuspended in 500 μl 2% formolsaline and stored at 4°C in the dark, prior to FACS analysis. Once optimised, experiments were all performed in triplicate.

3.2.12.2 p24 assay

To prepare plates, the antibody D7320 (Aalto Bio Reagents, Dublin, Ireland), diluted in coating buffer (100mM $NaHCO_3$, pH 8.5) was coated at 0.5 μg/well onto the inner 60 wells of Maxi-Sorb high-bind 96 well plates (NUNC, Roskilde, Denmark) by incubating them overnight, covered, at RT. The plates were then washed twice with 200 μl of TBS (144 mM NaCl, 25 mM Tris, pH 7.5), air dried before use or sealed and stored at -20°C. Samples were inactivated by adding Empigen zwitterionic detergent (Surfachem Ltd, Leeds, UK) to a final concentration of 1% by volume, followed optionally by incubation at 56°C for 30 minutes. The lysates were then stored at -20°C as required. The antigen capture works best at a final Empigen concentration of 0.05-0.1%, so serial dilutions were made in TBS / 0.05% Empigen and the volume of the wells adjusted to 100 μl with the same buffer. p24 standard is diluted initially into 1% Empigen (see below) before construction of a calibration curve in 0.1% Empigen (as for samples).

On the day of use, 200 μl/well TBS/2% Marvel fat free skimmed milk powder was added to the plates and incubated between 10 minutes – 1 h, washed once with 200 μl/well TBS and air-dried. Samples and standard (in duplicate) were added and incubated, the capture stage completed within 2-3 h at RT. Unbound p24 was washed away with 2 x 200 μl of TBS and bound p24 detected with pre-diluted EH12E1 – AP (MRC ADP452) at 0.5 μg/ml. This pre-dilution is at 1:2000 in TBS, containing 4% Marvel fat free skimmed milk powder, 20% sheep serum (DAKO code X503) and 0.5% Tween 20. Binding was complete after 60 minutes at RT.

Plates were then washed 5x with 200 µl/well TBS containing 0.5% Tween 20. To each well 100 µl Lumi-phos alkaline phosphatase (AP) substrate (Aureon Biosystems, Vienna, Austria) was added. This was incubated at RT, in the dark for 1 h. The plate was then read using a LUCY luminometer (Anthos Labtech Instruments, Austria), the detection range being from 30-1000 pg/ml p24.

3.2.12.3 Flow cytometry

Expression of cell surface CD4, CXCR4 and CCR5, and GFP expression, were determined by flow cytometry. Cells were taken from culture and washed twice in PBS with 0.01% azide. Cells were then incubated with CD4-PE, CCR5-PE and CXCR4-PE (section 3.2.11), in 20µl PBS containing 1% FBS and 0.1% azide per 10^6 cells for 1 hour at 4°C. Cells were also incubated with isotype control at the same concentration. Cells were washed 3 times in PBS with 0.01% azide. Ten thousand events were collected and analysed using a FACScan with Cellquest software (Becton Dickinson, UK).

3.2.12.4 Microscopy

CVG-37 cells were observed using an Axiovert 135 (Zeiss, Oberkocken, Germany) fluorescent microscope. Pictures were taken using a Contax 167 MT single lens reflex camera and E6 Ektachrome slide film (Kodak). Confocal images were generated by fluorescence microscopy (Axiovert 100 TV (Zeiss, Oberkocken, Germany) with a MRC 1024 Confocal (Bio-Rad, Hercules, CA)), captured using Lasersharp software (Lasersharp 2000, Bio-Rad) and edited using Confocal Assistant v4.02 software.

3.2.12.5 Drug sensitivity assay

CVG-37 cells were infected as described in section 3.2.12.1. Following spinoculation and washing with PBS, cells were resuspended in the appropriate complete medium and six aliquots of 100 µl added to separate wells of a 96 well plate (1 x 10^4 cells/well). To each well 100 µl of complete medium, containing the 2x the appropriate final dilution of antiretroviral drug, was added, and plates incubated at 37°C, for 5/7 days. The ARVs tested were Azidothymidine (AZT), Efavirenz (EFV), and Amprenavir (APV) (all gifts from Dr. D. Pillay), at 0.05, 0.01, 0.005, 0.0025 and 0.001µM. Plates were then sealed and centrifuged at 1200 rpm for 5 minutes and cell-free virus containing supernatants

removed and stored at –80°C. In the case of drug sensitivity assays for protease inhibitors, the virus containing supernatants were added to fresh CVG-37 cells (1 x 10^4 cells/well), in a two-step protocol (Pirounaki *et al.*, 2000). Infections were allowed to proceed for 7 days and then cells prepared for FACs as described (section 3.2.12.1). Once optimised experiments were all performed in triplicate.

3.3 Results

3.3.1 HIV primary isolate stocks

In order to characterise the effect of HIV genetic variation on growth phenotype, HIV-1 group M subtypes A, AE, B, C, D, F, group O and HIV-2 primary isolates were acquired. Limited basic information was provided (Table 3.0, section 3.1.3.1) so once low-passage stocks were produced, coreceptor usage was characterised and proviral DNA sequence verified.

3.3.1.1 Titration and coreceptor usage

A sample of each virus was obtained from the Central Facility for AIDS Reagents. This was passaged through PBMCs once in order to create a submaster stock (approximately 20 mls). 1 ml samples of the submaster stock were then expanded on five separate occasions to produce approximately 100 mls of virus (per isolate) with which titrations and subsequent experiments could be performed. As it was expected that the titre of the viruses produced would be particularly low (Moore and Ho, 1995), titrations were performed on NP2 CD4/CXCR4-positive and NP2 CD4/CCR5-positive cells. These cell lines were selected as they are highly susceptible to HIV infection when exposed to virus of the appropriate tropism (Soda *et al.*, 1999). The tropism for each virus, the titre of each virus stock and the mean titres calculated are summarised in Table 3.3. Individual stock titres were used to calculate experimental conditions. In the case of viruses that were found to be dual tropic, titre was equivalent on both NP2/CD4/CXCR4 and NP2/CD4/CCR5 cell lines. Accordingly, the mean of the titre on both cell lines is presented.

Table 3.3 Table of coreceptor usage, stock titres and mean titre, for each virus studied.

Subtype		Titre (ffu/ml^{-1})						Mean
	CoRe	SM	Wk1	Wk2	Wk3	Wk4	Wk5	titre^{+}
A	R5	$1x10^{2}$	*NR*	$2.2x10^{3}$	*NR*	*NR*	*NR*	$2.2x10^{3}$
AE	X4	$1x10^{2}$	*NR*	*NR*	*NR*	*NR*	*NR*	*NR*
B	X4/R5	$2.1x10^{5}$	$1.2x10^{6}$	$4.8x10^{5}$	$6.8x10^{5}$	ND	ND	$6.4x10^{5}$
C	R5	$2x10^{4}$	NR	$1.4x10^{3}$	$1.5x10^{3}$	*NR*	ND	$7.6x10^{3}$
D	X4/R5	$3x10^{3}$	$1.2x10^{3}$	$1.5x10^{3}$	$3.2x10^{3}$	$9.3x10^{3}$	ND	$3.6x10^{3}$
F	R5	$2.2x10^{2}$	$2.5x10^{2}$	$2.8x10^{3}$	$*8x10^{3}$	$*1x10^{3}$	ND	$*2.5x10^{3}$
O	X4/R5	$3.5x10^{2}$	$3x10^{3}$	$5.1x10^{3}$	$9.2x10^{3}$	$1.1x10^{4}$	$2.4x10^{3}$	$5.7x10^{3}$
ETP	X4/R5	$1.2x10^{4}$	$2.3x10^{4}$	$1.6x10^{3}$	ND	ND	ND	$1.2x10^{4}$

$^{+}$mean of titres calculated for all stocks of each isolate *this stock grew on both X4- and R5-NP2 cells. The titre is an average.

NR - no result, no infectious titre was detected in this stock

ND - not done, **CoRe** - coreceptor usage

Table 3.4 Table outlining clones of Gag, PR, and RT genes from which sequence was successfully generated.

Subtype	Gag	PR	RT
A	NC*	NC	✓
AE	NC	NC	✓
B	✓	✓	✓
C	✓	✓	✓
D	NC	✓	✓
F	✓	✓	✓
O	✓	✓	✓

*NC – not cloned

3.3.1.2 Sequencing and subtyping

To confirm the sequence integrity of viruses being studied genomic DNA was extracted from infected PBMCs, from which working stock virus was harvested. Degenerate primers were designed (Table 3.2, section 3.2.5) to amplify the *gag, PR* and *RT* genes, and PCR products were cloned into vectors for sequencing. The genes successfully cloned (as identified by colony screening PCR) are outlined in Table 3.4. Where sequencing was achieved a search was performed using BLAST and in all cases, where the sequence of the viral genes in question had been published, the corresponding reference sequence was matched. As sequencing was successful for the PR and/or RT genes of subtypes A, AE, B, C, D, F and group O, these sequences were subtyped in order to both confirm their subtype classification, and to test the fidelity of STAR. In all cases STAR assigned the correct subtype/type to each cloned PR or RT fragment.

3.3.2 Development of a reporter cell line

Several cell lines have been produced with the specific aims of both being sensitive to infection by X4- and R5-tropic strains of HIV, and the ability to 'report' infection in a readily detectable, rapid and antigen-independent manner. Specifically, studies of both neutralisation and drug sensitivity have been performed using HIV-1 susceptible cell lines with reporters including secreted alkaline phosphatase (Miyake *et al.*, 2003), luciferase (Spenlehauer *et al.*, 2001) and β-galactosidase (Pirounaki *et al.*, 2000). Not all such cell lines are T-cell in origin and some reporter systems require more experimental processing than others. One cell line was available, however, which

expressed humanised S65T green fluorescent protein (GFP) under control of the HIV-1 LTR: CEM-G (Gervaix *et al.*, 1997). This cell line exhibits low constitutive background fluorescence but a high level of GFP expression upon infection with HIV-1, over 2 to 4 days. Addition of inhibitors of reverse transcriptase and protease during HIV infection of this cell line has been shown to allow the accurate determination of drug susceptibility. Furthermore, quantitation of infectivity of viral preparations may be determined by assessment of number of cells infected in the first round of infection (Gervaix *et al.*, 1997). The CEM-G reporter cell line, however, was designed for study of syncitium-inducing (X4-tropic) HIV-1 strains only. CEM cells do not endogenously express CCR5, so this cell line was chosen for modification such that it could be used to directly assess the *in vitro* growth characteristics of HIV primary isolates of differing subtype and cell tropism (Table 3.3, section 3.3.1.1).

3.3.2.1 Creating a CCR5+ CEM-G cell line

CCR5 was introduced to the CEM-G cell line by infection at high multiplicity with the VSV-G pseudotyped MLV-based retroviral vector pBABE-puro.CCR5 (Morgenstern & Land, 1990). Because of the reported high efficiency of both infection and expression using this system, once selected using puromycin, the bulk cell population was analysed for CCR5 expression (Figure 3.0-A). This revealed that, despite a small number of cells expressing CCR5 the majority of the population, whilst puro-resistant, did not express detectable CCR5. Accordingly, limiting dilutions of the bulk puro-resistant cell population were prepared and clonal lines developed. Forty clones were selected and expanded further under selection, each then stained with an isotype control antibody and anti-CCR5-PE, to determine whether CCR5 was expressed on the cell surface. The results for all clones are shown in Figure 1, Appendix. It was expected that, given the results from staining and FACs analysis of the bulk population, a few of these clonal lines would express high levels of CCR5. The FACS data for the forty clonal lines revealed that 14 cell lines appeared to express CCR5. These, therefore, were expanded further and the expression level of CD4, CXCR4 and CCR5 determined. The results are shown in Figure 3.0-B; six clonal lines expressed a low/medium level of CCR5, Figure 3.0-C; three clonal lines expressed a medium level of CCR5, and Figure 3.0-D; two clonal lines expressed the highest level of CCR5. PBMCs were simultaneously stained for CD4, CXCR4 and CCR5 (Figure 3.0-E) and expression profile of these receptors was remarkably similar to that seen in CEM-G.R5 clones 33 and 37.

Figure 3.0 FACs data for development of CCR5 expressing CEM-G cell lines.
(A) FACs data showing expression of CD4, CXCR4 and CCR5 on bulk population of puromycin selected CEM-G cells infected with pBabe-puro.CCR5 retroviral vector. Whilst results show that there is no bulk expression of CCR5, certain puro-resistant CCR5 positive cells are present (marked with arrows), **(B)** FACs data showing expression of CD4, CXCR4 and CCR5 on clonal lines, produced by limiting dilution under puromycin selection of CEM-G cells infected with pBabe-puro.CCR5. Clones 7, 34, 39, 22, 14 and 28 are low/medium expressors of CCR5.

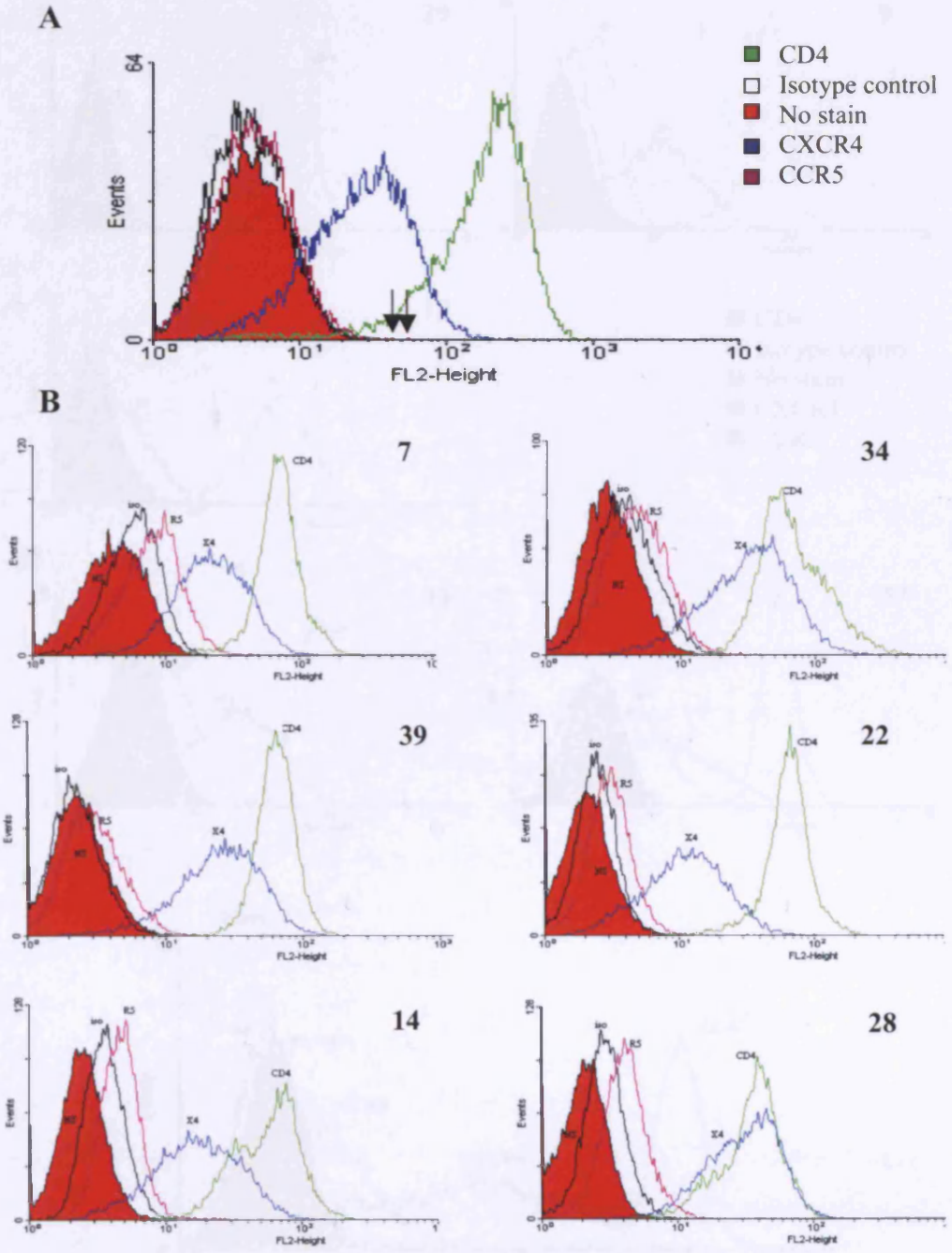
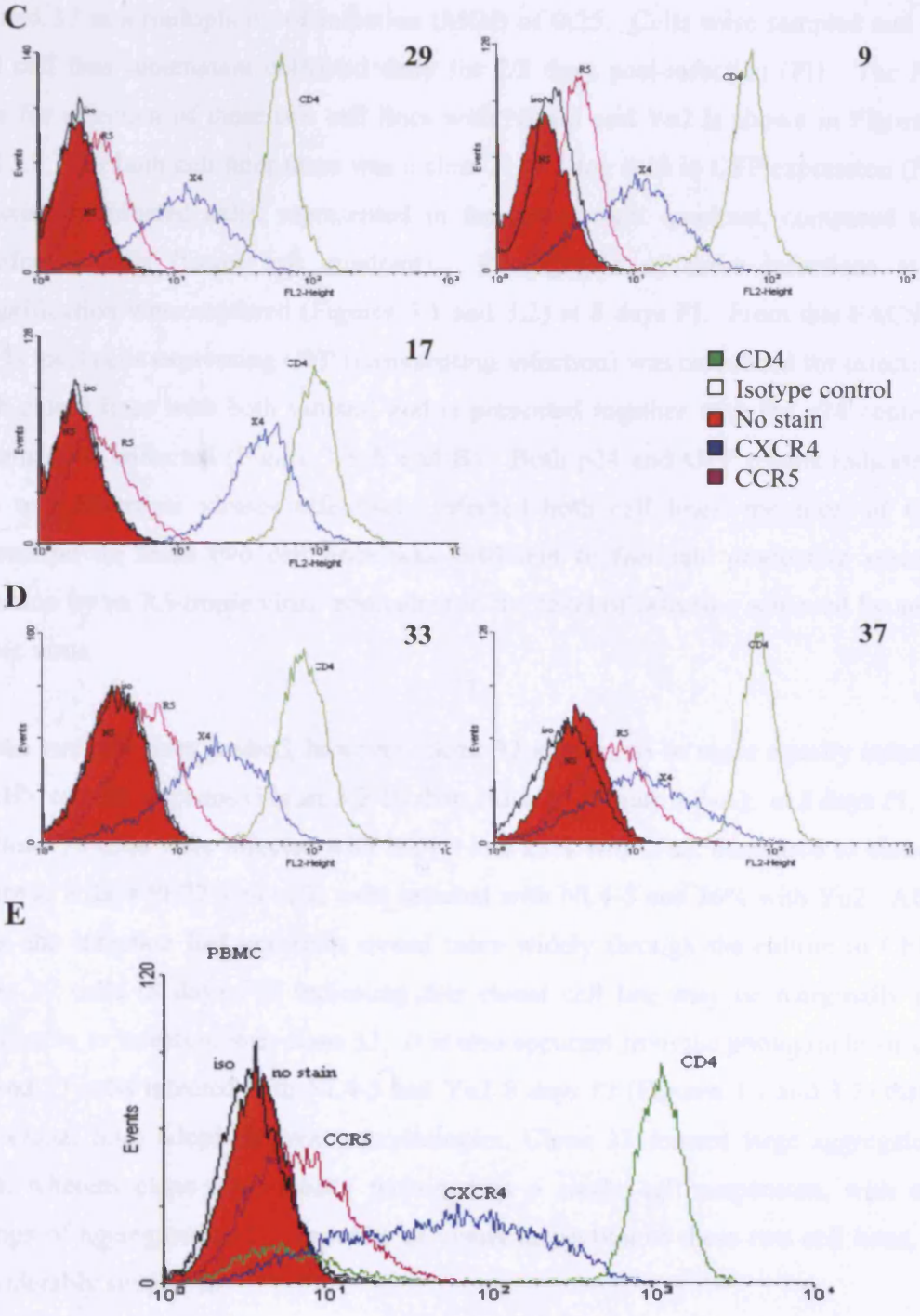


Figure 3.0 (continued) FACs data showing expression of CD4, CXCR4 and CCR5 on clonal lines, produced by limiting dilution under puromycin selection of CEM-G cells infected with pBabe-puro.CCR5.
(C) Clones 29, 9 and 17 express medium levels of CCR5, and **(D)** Clones 33 and 37 express the highest level of CCR5. The profile of these cells in terms of receptor expression is similar to **(E)** PBMCs



3.3.2.2 Clones 33 and 37 can be infected with both X4- and R5-tropic viruses

Given identification of two clonal cell lines that expressed the CCR5 coreceptor to a level comparable to that detected on PBMCs, these were infected with the X4- and R5-tropic viruses NL4-3 and Yu2. These are infectious molecular clone derived viruses, high titre stocks of which were generated to enable infection of the CEM-G.R5 clones 33 and 37 at a multiplicity of infection (MOI) of 0.25. Cells were sampled and fixed and cell free supernatant collected daily for 7/8 days post-infection (PI). The FACS data for infection of these two cell lines with NL4-3 and Yu2 is shown in Figures 3.1 and 3.2. For both cell lines there was a clear 2- to 3-log shift in GFP expression (FL1 – X axis) in infected cells, represented in the lower right quadrant, compared to the uninfected cells (lower left quadrant). Photographs of these infections at low magnification were captured (Figures 3.1 and 3.2) at 8 days PI. From this FACS data the % total cells expressing GFP (representing infection) was calculated for infection of both clonal lines with both viruses, and is presented together with the p24 content of supernatants collected (Figure 3.3-A and B). Both p24 and GFP results indicate that X4- and R5-tropic viruses effectively infected both cell lines: the level of CCR5 expression on these two cell lines was sufficient to facilitate productive spreading infection by an R5-tropic virus, equivalent to the level of infection achieved by an X4-tropic virus.

Of the two cell lines studied, however, clone 37 appears to be more equally infectable by HIV of both tropisms (Figure 3.3-B) than clone 33 (Figure 3.3-A): at 8 days PI, 13% of clone 33 cells were infected with NL4-3 and 23% with Yu2, compared to clone 37, where at 8 days PI 27% of cells were infected with NL4-3 and 26% with Yu2. After 8 days, the infection had generally spread more widely through the culture in CEM-G clone 37 cells (8 days PI), indicating this clonal cell line may be marginally more permissive to infection than clone 33. It is also apparent from the photographs of clone 33 and 37 cells infected with NL4-3 and Yu2 8 days PI (Figures 3.1 and 3.2) that the two clonal lines adopt different morphologies. Clone 33 formed large aggregates of cells, whereas clone 37 typically remained as a single cell suspension, with small clumps of aggregated cells. In terms of visual inspection of these two cell lines, it is considerably simpler to

Figure 3.1 **FACS data for an infection time course of HIV-1 NL4-3 (X4-tropic) and Yu2 (R5-tropic), on CEM-G.R5 clone 33 cells.**
The increasing number of cells in the lower right quadrant, over time, represents the increasing number of cells in which GFP is expressed and are thus infected. The inset pictures were taken at 8 days PI, and demonstrate the appearance of both infected and uninfected clone 33 cells.

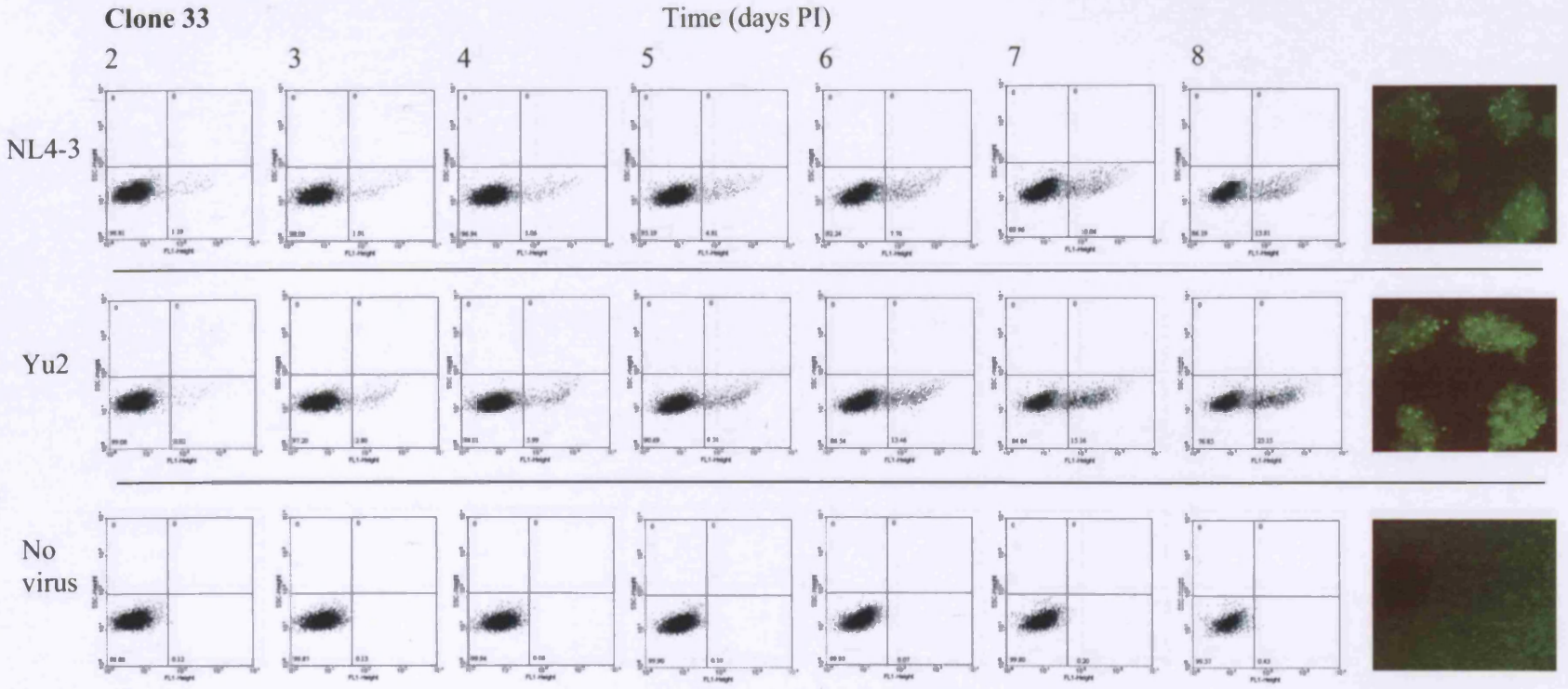


Figure 3.2 **FACS data for an infection time course of HIV-1 NL4-3 (X4-tropic) and Yu2 (R5-tropic), on CEM-G.R5 clone 37 cells.**
The increasing number of cells in the lower right quadrant, over time, represents the increasing number of cells in which GFP is expressed and are thus infected. The inset pictures were taken at 8 days PI, and demonstrate the appearance of both infected and uninfected clone 37 cells.

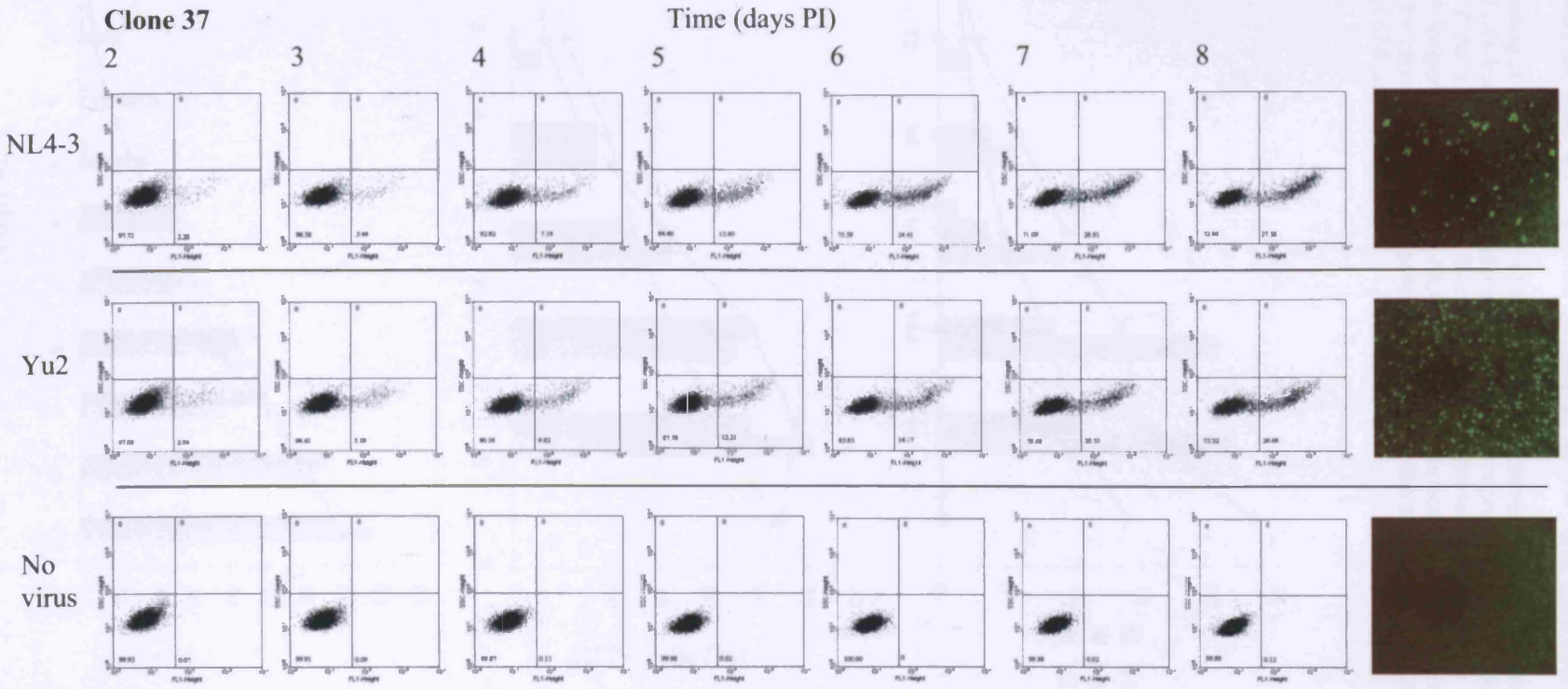
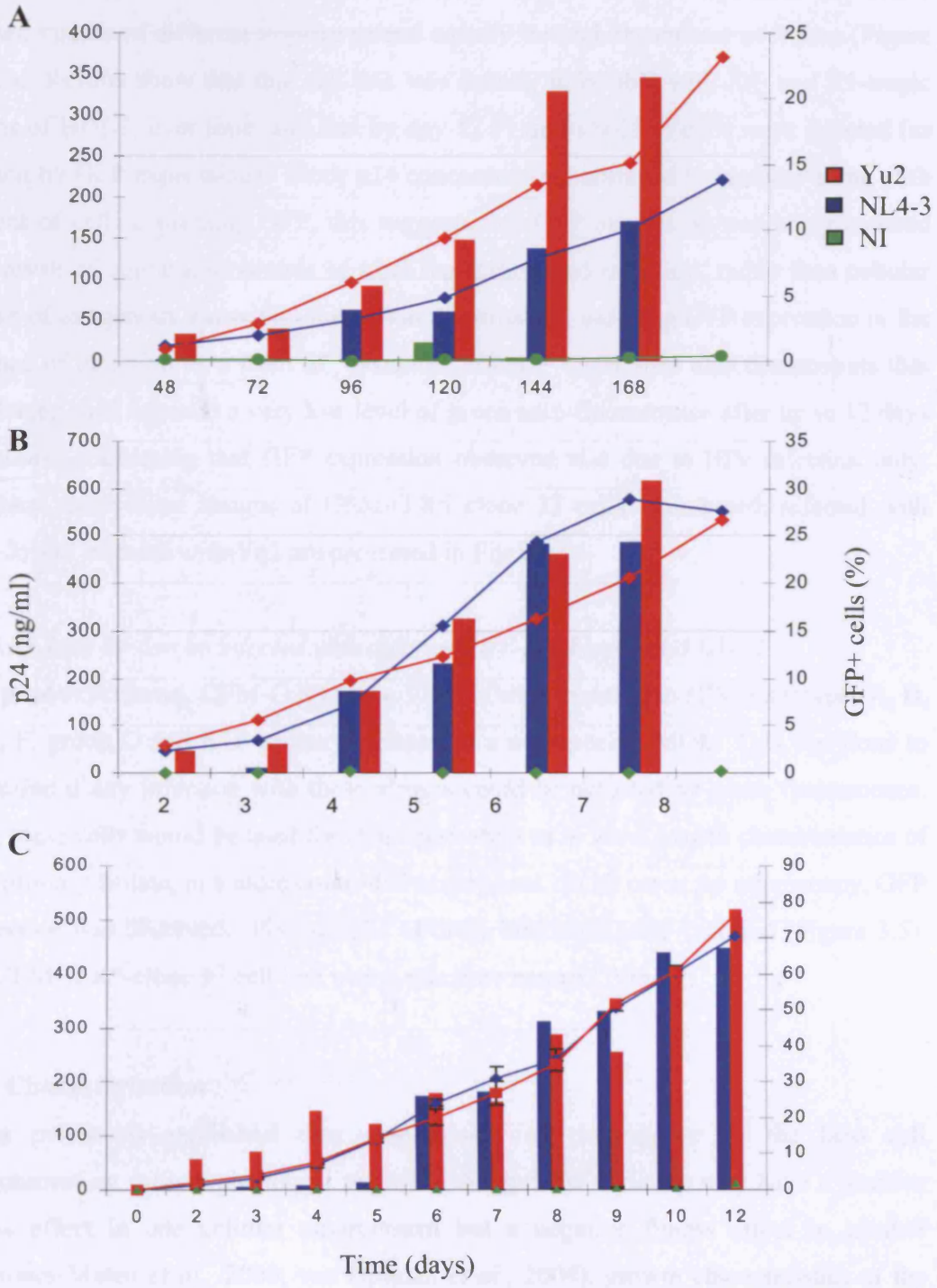


Figure 3.3 Susceptibility of CEM-G.R5 clones 33 and 37 to X4- and R5-tropic HIV infection.

Graphs representing the testing of permissiveness of CEM-G.R5 **(A)** clone 33 and **(B)** clone 37, to infection by HIV-1 NL4-3 (X4-tropic) and Yu2 (R5-tropic). The bars represent p24 detected in cell-free supernatant over 7 days, and the lines % of total cells expressing GFP. This data extends to 8 days PI. **(C)** A longer time course of NL4-3 and Yu2 infection of CEM-G clone 37 was performed, to assess the reproducibility of infection and how it spreads through the culture. GFP data is in triplicate and p24 duplicate.



interpret the level of infection when looking at clone 37, rather than clone 33: the internal reflection of green fluorescence within the aggregates of infected and un-infected clone 33 cells creates a deceptively high GFP signal, as interpreted by eye.

As clone 37 seemed preferable, the infection of this cell line with NL4-3 and Yu2 was repeated in triplicate. Cells and supernatant were sampled for 12 days PI to establish if the two viruses of different tropism spread equally through the culture over time (Figure 3.3-C). Results show that this cell line was equally infectable with X4- and R5-tropic strains of HIV-1, over time, and that by day 12 PI up to 80% of cells were infected (as defined by GFP expression). Since p24 concentration continued to increase along with percent of cells expressing GFP, this suggests that GFP expression was being induced as a result of continuous rounds of viral replication and infection, rather than cellular uptake of exogenous Tat within the culture supernatant, inducing GFP expression in the absence of infection as a form of 'bystander effect'. These data also demonstrate that uninfected cells retained a very low level of green auto-fluorescence after up to 12 days in culture, confirming that GFP expression observed was due to HIV infection only. Confocal microscope images of CEM-G.R5 clone 37 cells: uninfected; infected with NL4-3; and infected with Yu2 are presented in Figure 3.4.

3.3.2.3 Clone 37 can be infected with different HIV-1 subtypes and HIV-2

As a pilot experiment, CEM-G.R5 clone 37 cells were exposed to HIV-1 subtypes A, B, C, D, F, group O and ETP primary isolates, at a non-specific MOI. This was done to determine if any infection with these viruses could be detected by green fluorescence. If so, these cells would be used for characterisation of *in vitro* growth characteristics of each primary isolate, in a more controlled experiment. In all cases, by microscopy, GFP expression was observed. Photographs of these infections were captured (Figure 3.5). The CEM-G.R5-clone 37 cell line was at this time named CVG-37.

3.3.3 Characterisation

Given previously published data emphasising the importance of the host cell environment on growth phenotype *in vitro* - that genetic variation may have a positive fitness effect in one cellular environment but a negative fitness effect in another (Quinones-Mateu *et al.*, 2000, van Opijnen *et al.*, 2004), growth characteristics of the

Figure 3.4 Phase contrast and green fluorescent panels showing CEM-G.R5 clone 37 cells 4 days PI.
(A) Cells are infected with the X4-tropic HIV-1 strain NL4-3, **(B)** the R5-tropic strain Yu2, and **(C)** uninfected cells.

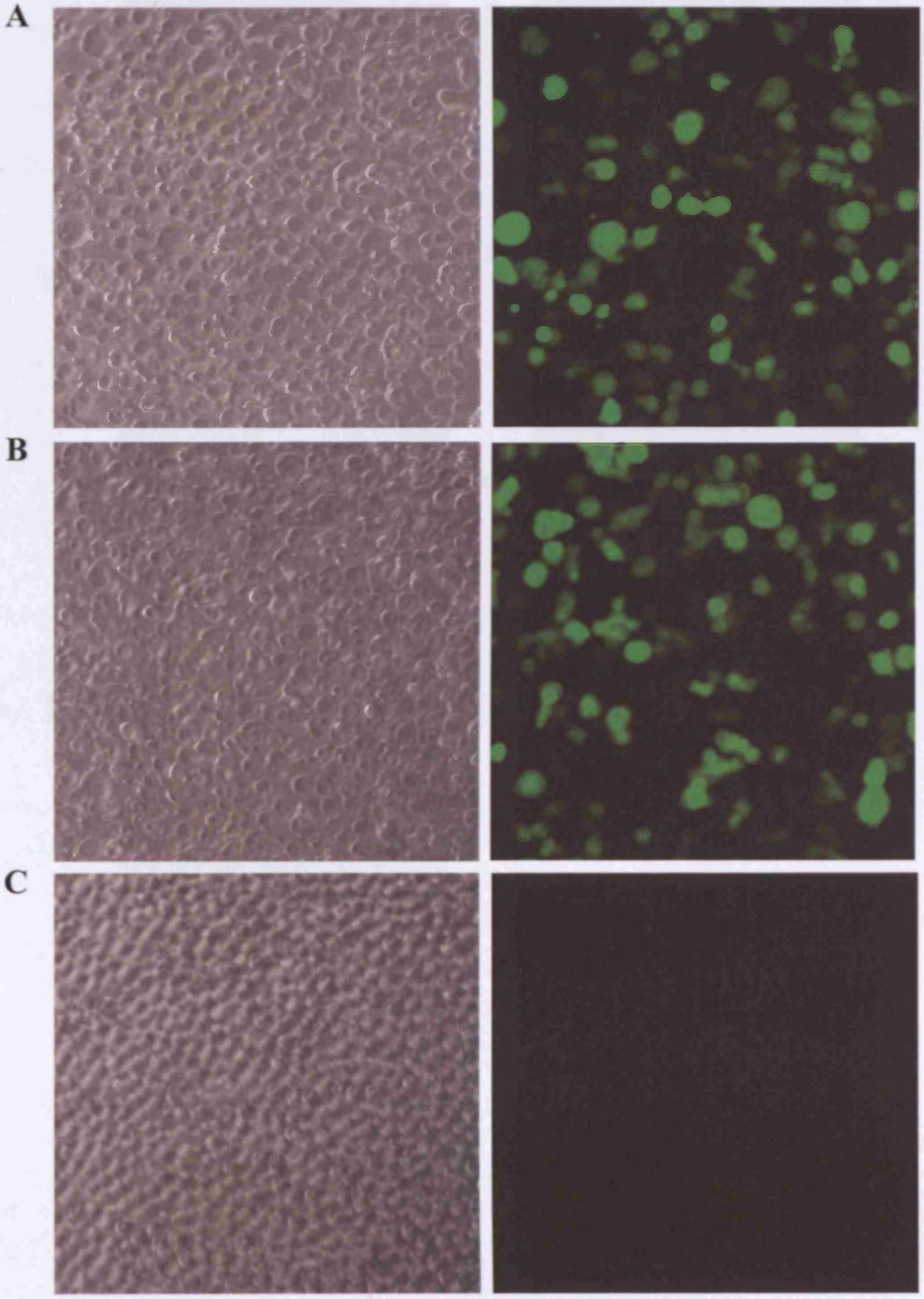
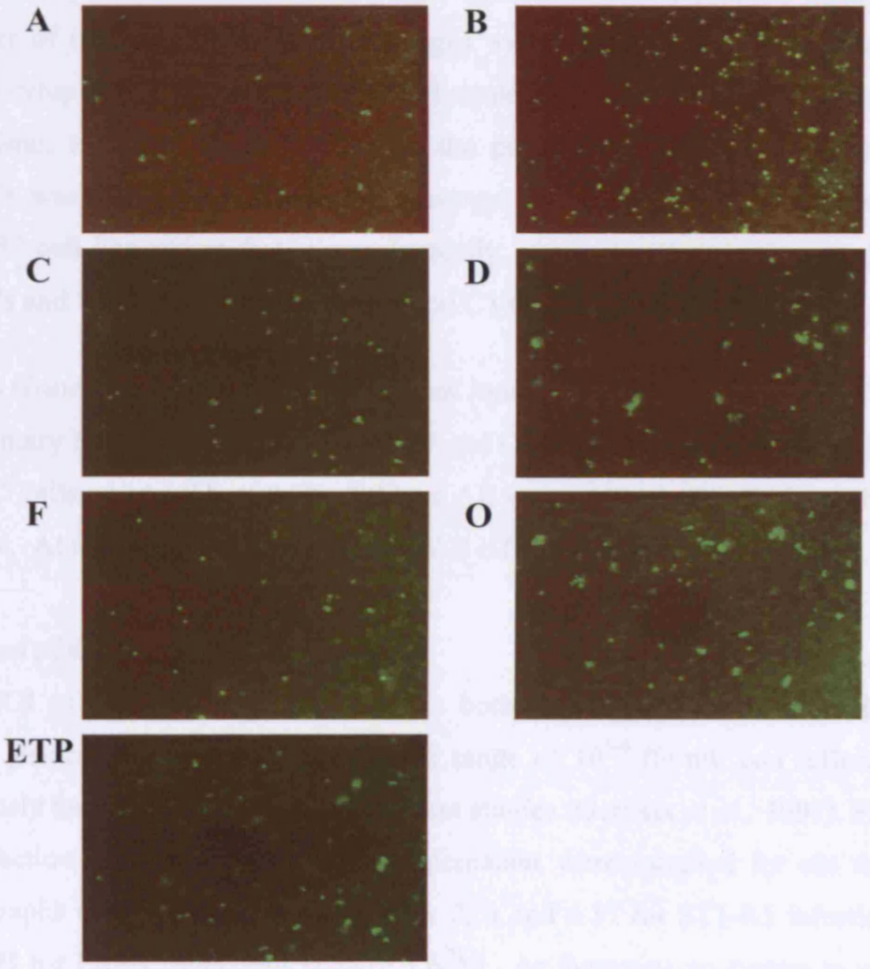


Figure 3.5 Photographs of CEM-G.R5 clone 37 cells, infected with HIV-1 subtype A, B, C, D, F, Group O and HIV-2 ETP primary isolate viruses, under low power magnification. Images were captured using a fluorescent microscope and single lens reflex (SLR) camera. Cells expressing GFP are evident in all infections, although some more than others. This experiment was not performed at equalised MOI so does not infer any biological comparisons between isolates.



panel of HIV-1 primary isolates and HIV-2 were assessed in two different cellular environments. Specifically, the CVG-37 reporter cell line was used for infection studies, with growth assessed by both p24 and green fluorescence (FACS and microscopy). In addition, a SupT1 T-cell line that has been engineered to express CCR5 (ST1-R5) was used for comparative infection studies, but as there is no direct reporter of infection in these cells, images were captured using light microscopy (to record cytopathic effects – CPE) and p24 content of cell-free supernatant was recorded over time, to give a representation of the progress of infection. Virus growth on PBMCs was also captured using microscopy and photography. Development of the CVG-37 cell line meant that, chronologically, virus growth was characterised first on PBMCs and ST1-R5s, and subsequently on CVG-37.

3.3.3.1 Growth curves on PBMC and T-cell lines

Six primary HIV-1 isolates; A, B, C, D, F and O, were used for infection of PBMCs and ST1-R5 cells, at an MOI of 0.01. Subtype AE was not used, as infectious viral titre was too low. At this stage titred stocks of HIV-2 ETP had not yet been acquired.

PCR and p24

The MOI of 0.01 was chosen, as this is both realistically achievable using primary isolate stocks with infectious titre in the range of 10^{3-4} ffu/ml, and reflects the titres previously used for *in vitro* viral replication studies (Gervaix *et al.*, 1997). For the ST1-R5 infection time courses cells and supernatant were sampled for six days PI and photographs of cultures captured on days 2, 4 and 6 PI for ST1-R5 infections, days 1 and 3 PI for PBMC infections (Figure 3.6-A). As there was no system to immediately report infection (as in CVG-37 cells) and it was possible that the p24 assay may not be optimal for certain HIV isolates (in particular those which are more distant to HIV-1 subtype B, for which the assay was optimised), PCR for the early reverse transcription product strong stop (SS) DNA (Li *et al.*, 1993) was employed to confirm that active HIV-1 replication was responsible for any changes in CPE detected. Specifically, DNA was extracted from ST1-R5 cells infected with each virus, at each time point, and the success of DNA extraction determined by performing PCR to amplify a highly conserved region of the human endogenous retrovirus (ERV3) DNA, found within the genome of every cell (O'Connell *et al.*, 1984). It was expected that DNA extraction would be difficult as only 10^4 cells were infected for each time point, so screening for

ERV3 enabled the selection of a time point from which DNA was extracted for all samples to screen for SS-DNA, thus avoiding false negative results (Figure 3.6-C1). 96 hours post-infection (HPI) was chosen as all samples were ERV3 positive and screening of these revealed that in all cases except the group O virus infection, SS-DNA could be detected (Figure 3.6-C2), confirming ongoing HIV-1 replication.

p24 data for the infection of ST1-R5 cells is shown in Figure 3.6-B. Results show that of all viruses, subtype D appeared to replicate the most efficiently in ST1-R5 cells – p24 being detectable 1 day PI and accumulating in cell free supernatants up to day 6 PI when the concentration of p24 detected is over twice that seen in the next most efficient viruses, HIV-1 subtype A and group O. In the case of subtype A, however, infection (as represented by p24 concentration) appeared to peak 4 days PI, whereas p24 in the group O infection did not appear until 4 days PI, and was still increasing at 6 days PI. Subtype C infection followed a similar pattern to subtype A, and subtype F was similar to group O in that p24 appeared late, although to a much lower level. p24, however, was not detected in subtype B infection, implying that the virus used was non-infectious, or that subtype B was restricted in ST1-R5 cells. The latter possibility is, however, unlikely as SS-DNA was detected in this time course at 96 HPI. It is also unlikely that the p24 assay failed, as it was designed with subtype B capsid in mind. These experiments were repeated in triplicate, however, and the patterns of differential viral replication were consistent.

Microscopy

Photographic images of these time courses (for which PCR and p24 data have been described) taken using a single lens reflex camera and light microscope are also presented, showing the CPE evident at days 2, 4 and 6 PI for ST1-R5 infections, and days 1 and 3 PI for PBMC infections. Figure 3.7 shows uninfected ST1-R5 and PBMCs, cultured in the same conditions as the infected cells, over 6 days. By comparing these to images those of both infected ST1-R5 cells (Figure 3.8-A and B) and PBMCs (Figure 3.9-A and B) it is clear that, by microscopy, infection can be confirmed in both cell types as a result of exposure to all viruses. The X4- or dual-tropic viruses HIV subtype B, D and O cause a heavy burden of CPE in infected ST1-R5 cultures by 96 HPI, largely in the form of

Figure 3.6 **HIV-1 subtype growth comparison experiments on ST1-R5 cells**
(A) Schematic of comparative growth experiment protocol, **(B)** p24 concentration in cell-free culture supernatant, measured over 6 days, for cells infected with HIV-1 subtypes A, B, C, D, F and O, **(C-1)** Erv3 PCR from DNA extracted from cells at each time point. At 96 hours Erv3 PCR indicates successful DNA extraction. Accordingly, this DNA was used for HIV strong-stop PCR, shown in **(C-2)**. Strong-stop DNA was detected in all infections, except for with group O virus.

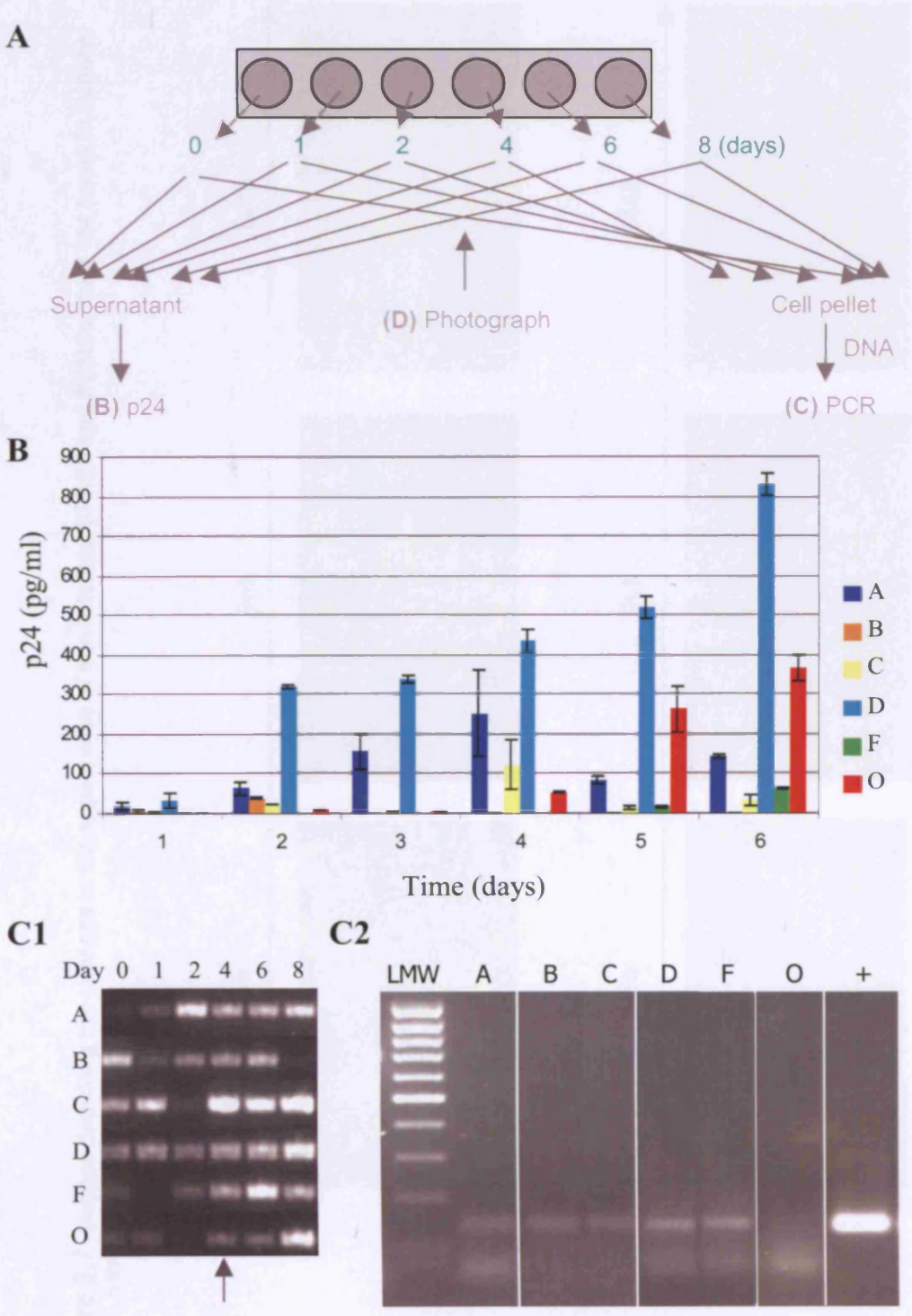


Figure 3.7 **Images depicting the changes in the appearance of non-infected ST1-R5 and PBMCs, over 144 hours in culture.**
The 96 h image for PBMCs is missing.

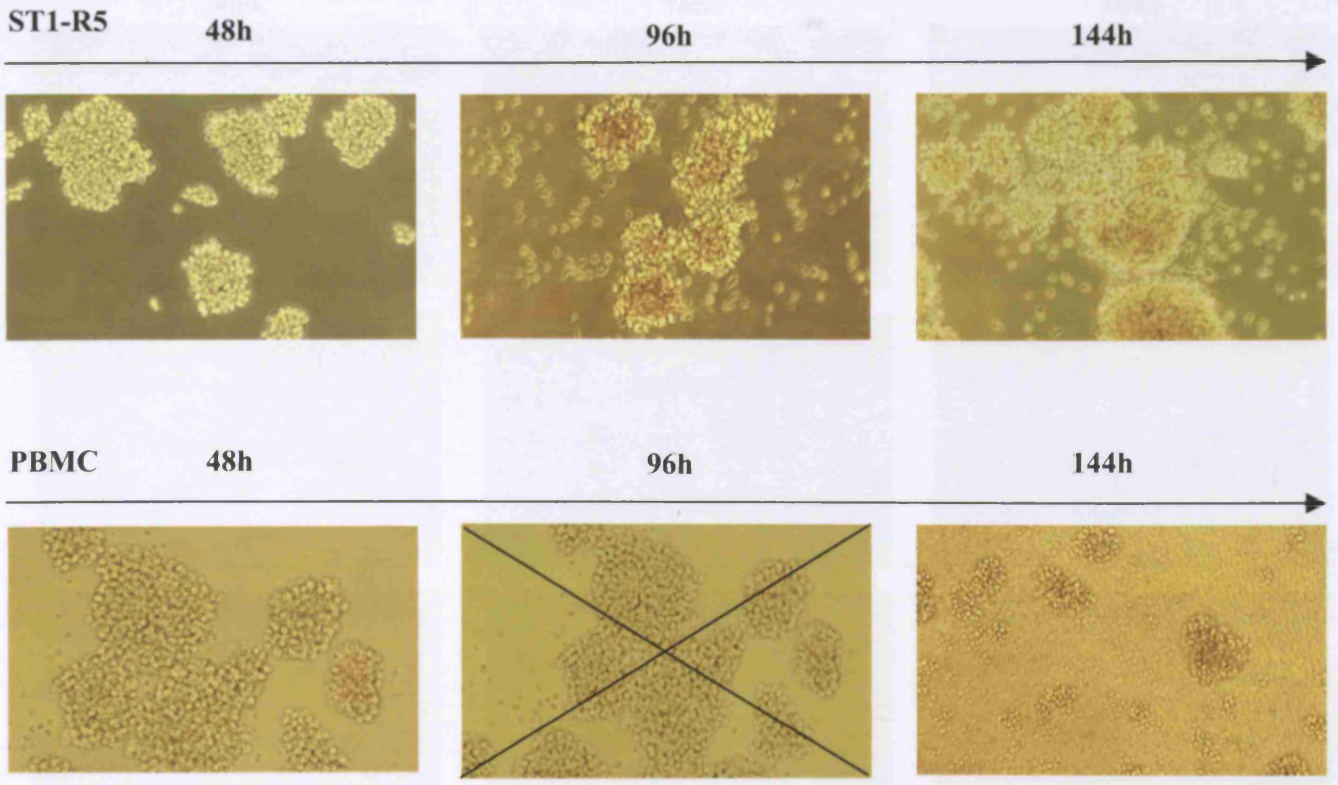


Figure 3.8 - a. **Images depicting the changes in the appearance of ST1-R5 cells, over 144 hours in culture, infected with HIV-1 primary isolates of subtype A, B and C.**

Coreceptor usage of each isolate is noted on the right of each panel of images. Arrows mark large multinucleate cells/syncitia characteristic of infection with SI or X4-using viruses.

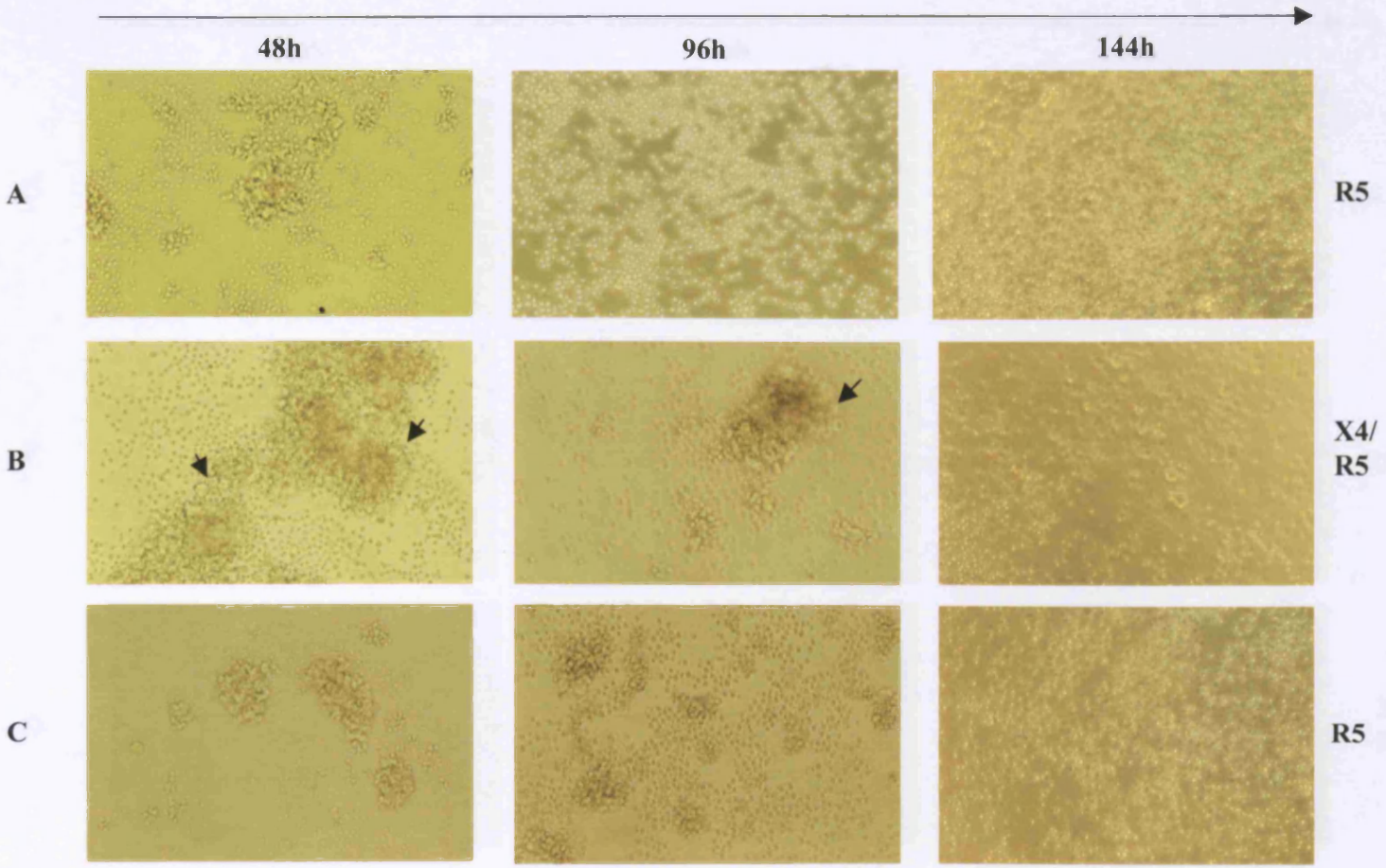


Figure 3.8 – b. **Images depicting the changes in the appearance of ST1-R5 cells, over 144 hours in culture, infected with HIV-1 primary isolates of subtype D, F and O.**

Coreceptor usage of each isolate is noted on the right of each panel of images. Arrows mark large multinucleate cells/syncitia characteristic of infection with SI or X4-using viruses.

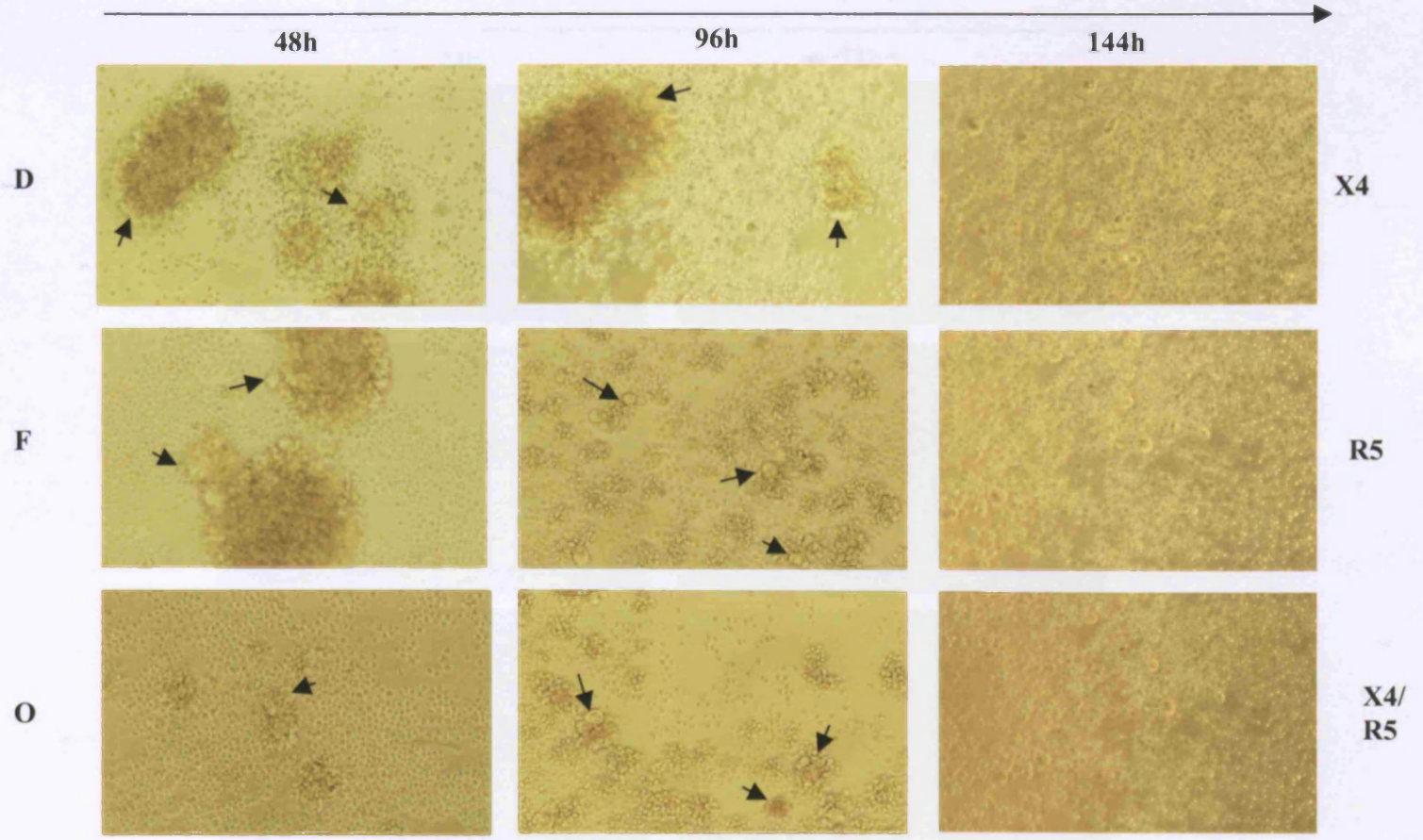


Figure 3.9 – a. Images depicting the changes in the appearance of PBMCs, over 144 hours in culture, infected with HIV-1 primary isolates of subtype A, B and C.

Coreceptor usage of each isolate is noted on the right of each panel of images. Arrows mark large multinucleate cells/syncitia characteristic of infection with SI or X4-using viruses.

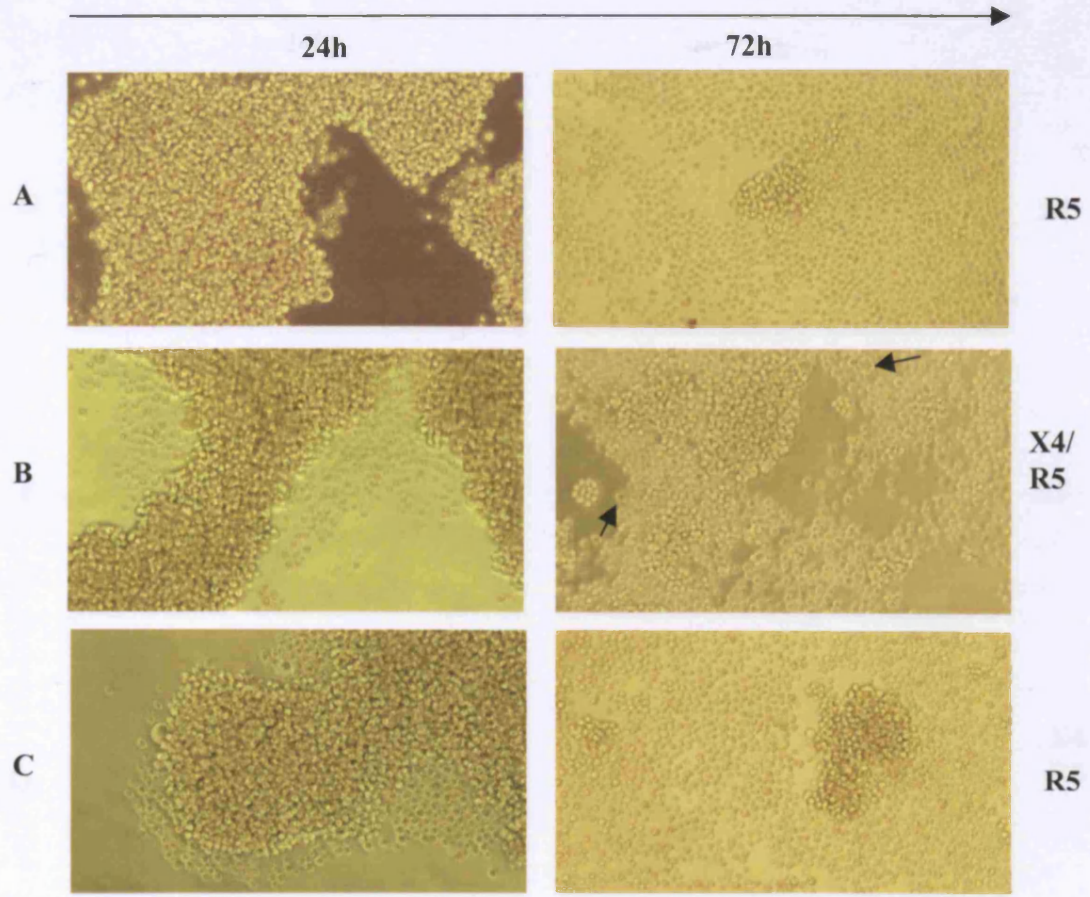
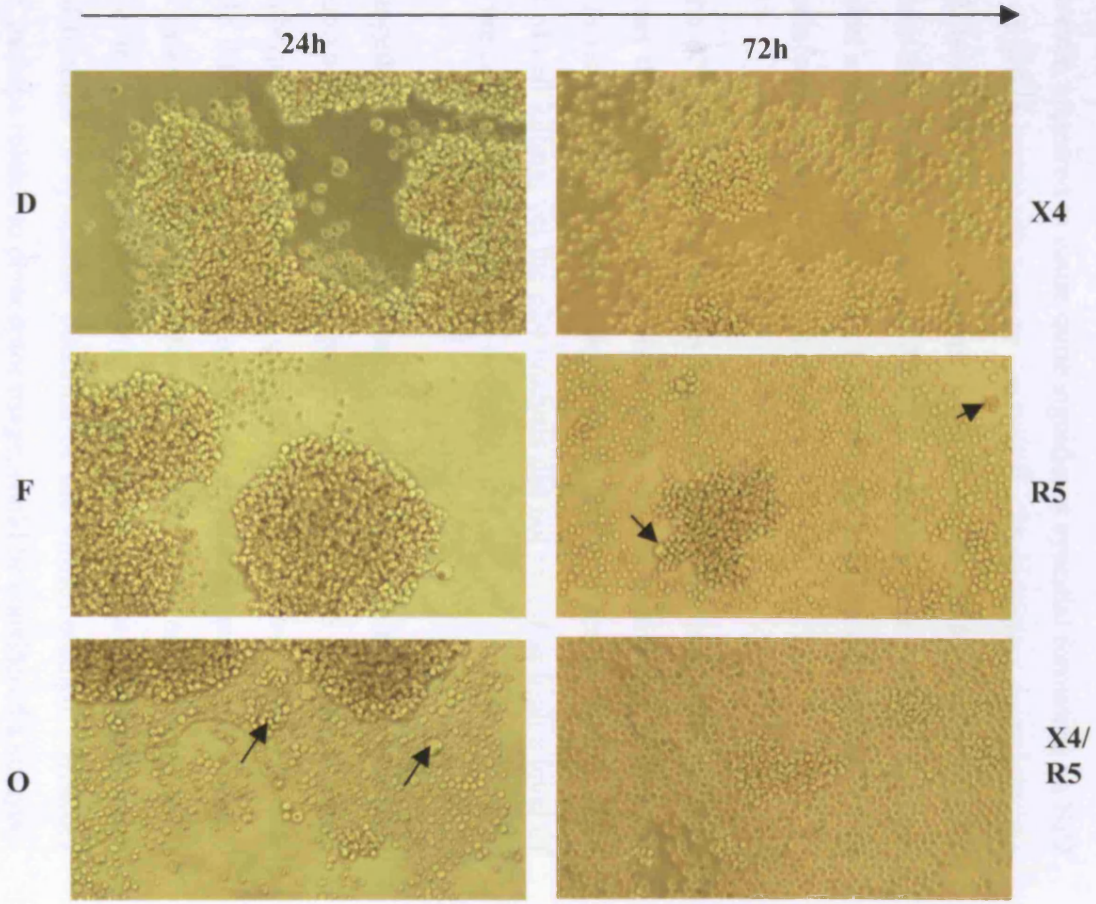


Figure 3.9 – b. Images depicting the changes in the appearance of PBMCs, over 144 hours in culture, infected with HIV-1 primary isolates of subtype D, F and O.

Coreceptor usage of each isolate is noted on the right of each panel of images. Arrows mark large multinucleate cells/syncitia characteristic of infection with SI or X4-using viruses.



syncitia and multinucleate cell formation (marked by arrows). This is typical of X4-tropic, or SI/'fast' viruses, as first characterised on MT2 cells (Connor *et al.*, 1997), and in all cases by 96 HPI cell aggregates had broken down and the majority of cells in each field of view appeared to have lost biofringence and be either dead or undergoing apoptosis. The R5-tropic viruses HIV subtype A and C also held to this dogma, in that they failed to induce syncitia yet still caused pyknosis and apparent degradation of individual cells, leading to a similar end point in terms of observable CPE, by 144 HPI. HIV subtype F, however, appeared to cause quite significant syncitial formation on NP2 cells. The images of infected PBMCs largely confirm the observations made with the ST1-R5 cells. CPE, however, was less clear as PBMCs tend to not form aggregates in culture so the blister-like syncitia clear in ST1-R5 infection were not so evident. Trypan blue exclusion analysis of these cells, however, did confirm cell killing at 144 HPI, compared to uninfected controls (data not shown).

When comparing the p24 data and the images captured during the infections yielding these results, it seems that the p24 data in some cases do not necessarily reflect the extent of infection *in vitro*. For example, subtype B, C, and F infections were clearly productive in terms of cell killing, yet the p24 analysis did not reveal as high a level of this viral protein in the cell-free supernatant as one might expect.

When considering together the p24 data and microscopic analysis of the infection time courses of HIV-1 subtypes A, B, C, D, F and group O, on the ST1-R5 cell line, it seems that the different subtypes of virus do grow quite differently *in vitro*. The p24 data showed reproducible differences in the rate of p24 accumulation in the supernatant. This seems in some cases to agree with the observed CPE (i.e. in the case of subtype D) but not in others (i.e. in the case of subtype F). It may be that these differences were purely a result of difference in cytopathic potential of the viruses studied. Whether cytopathic potential, perhaps related to coreceptor usage, could be considered a subtype-specific trait, however, is a controversial point. It was hoped that by performing the infection time courses with viruses of different subtype using CVG-37 cells, this would provide a clearer representation of the apparent differences in replication capacity, between HIV-1 genetic variants.

3.3.3.2 Growth curves on CVG-37

With the CVG-37 reporter cell line, infections were also performed at a multiplicity of infection of 0.01 ffu/cell. In this case, however, cells were infected by spinoculation followed by incubation in the presence of polybrene, in an attempt to increase the infection efficiency for all viruses. p24 was again measured for each time point and each virus, and the extent of infection expressed as percentage of total cells expressing GFP (measured using FACS).

Results from analysis of GFP expression show that the HIV-1 subtypes A, D, F, group O and HIV-2 ETP spread through the culture equally, all achieving approximately 10-15 % infection 5 days PI (Figure 3.10-B). From day 5 onwards, however, the pattern of spread of these viruses diverged, with HIV-1 subtype D and group O continuing to spread resulting in 60% infection by 12 days PI. HIV-2 ETP also spread but to a lesser extent, achieving 30-35% infection, but HIV-1 subtypes A and F appeared to decline and fail to infect more that 10-15% of the CVG-37 cells. Similar to the growth experiments on ST1-R5 cells, HIV-1 subtypes B and C failed to establish significant levels of infection, only 5% of cells expressing GFP as a result of infection, over the 12 day time course. The FACS plots for the 12 days PI time point are shown in Fig 3.11: GFP expressing infected cells can be clearly delineated from non-infected GFP negative cells. Furthermore, the level of shift along the X-axis is comparable to that achieved by high titre infection with tissue culture adapted virus. This therefore shows that the use of low titre primary isolate viruses does not adversely affect the capacity of viral infection to induce GFP expression.

p24 assays performed using the cell–free supernatants from the infected CVG-37 cells assessed for GFP expression largely confirm the observations from GFP data (Fig 3.10-A). Specifically, HIV-1 group O and subtype D p24 levels are highest at day 12, corresponding to the level of infection detected by GFP expression. p24 was also detected in subtype F, A and B infections, but only to a low level and not necessarily corresponding to when virus replication was detectable by GFP expression. This again highlights a limitation of the p24 assay in that it was insufficiently sensitive to detect the progress of infection within a culture, over time. The levels of p24 detected in this series of experiments, however, is much greater than that detected in the ST1-R5

Figure 3.10 **HIV-1 subtype growth comparison experiments in the CVG-37 reporter cell line.**
(A) p24 concentration in cell-free culture supernatant, and **(B)** percentage of cells expressing GFP, both measured over 12 days, as a result of HIV infection of CVG-37 cells with HIV-1 subtypes A, B, C, D, F, group O and HIV-2 ETP.

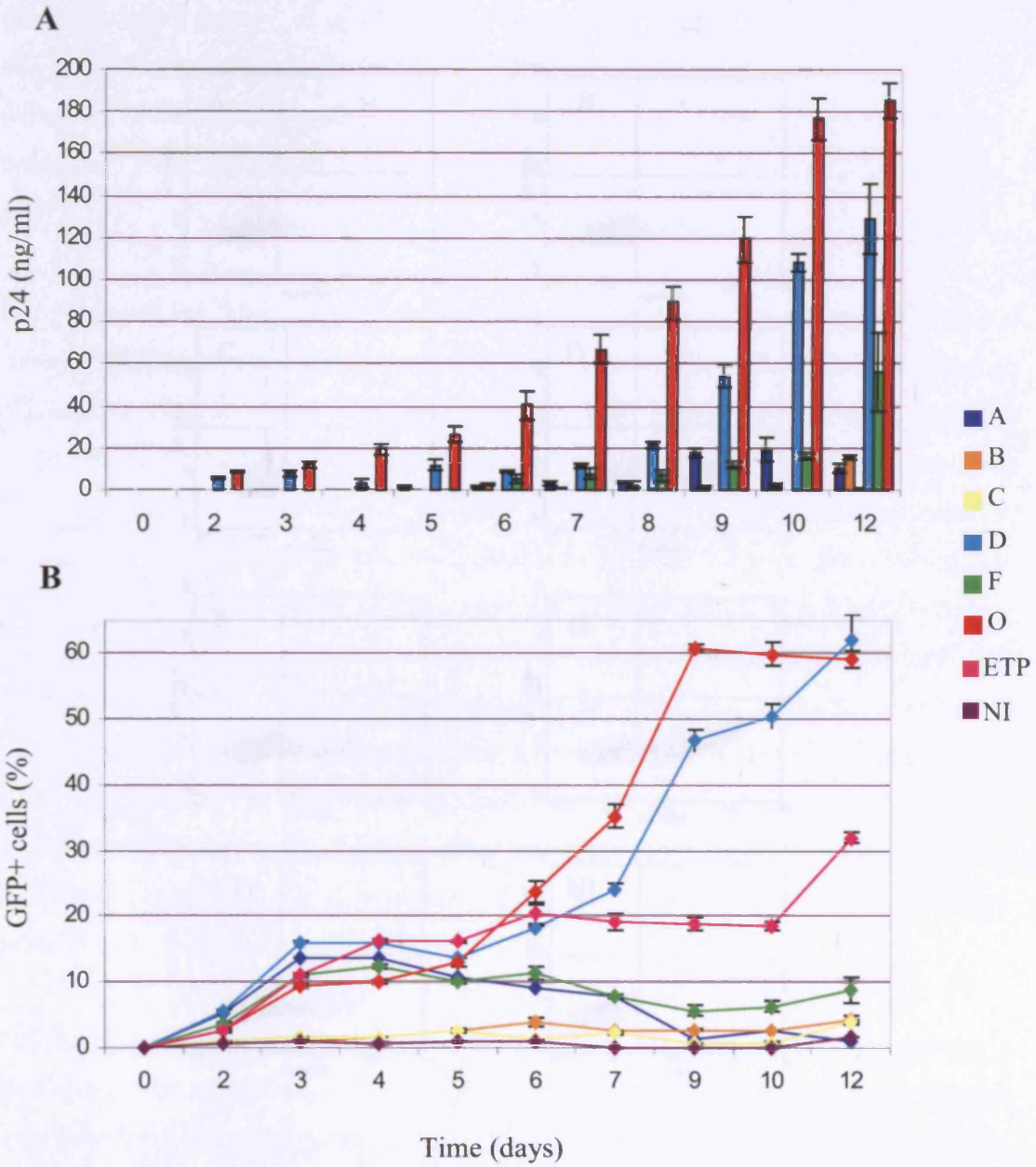
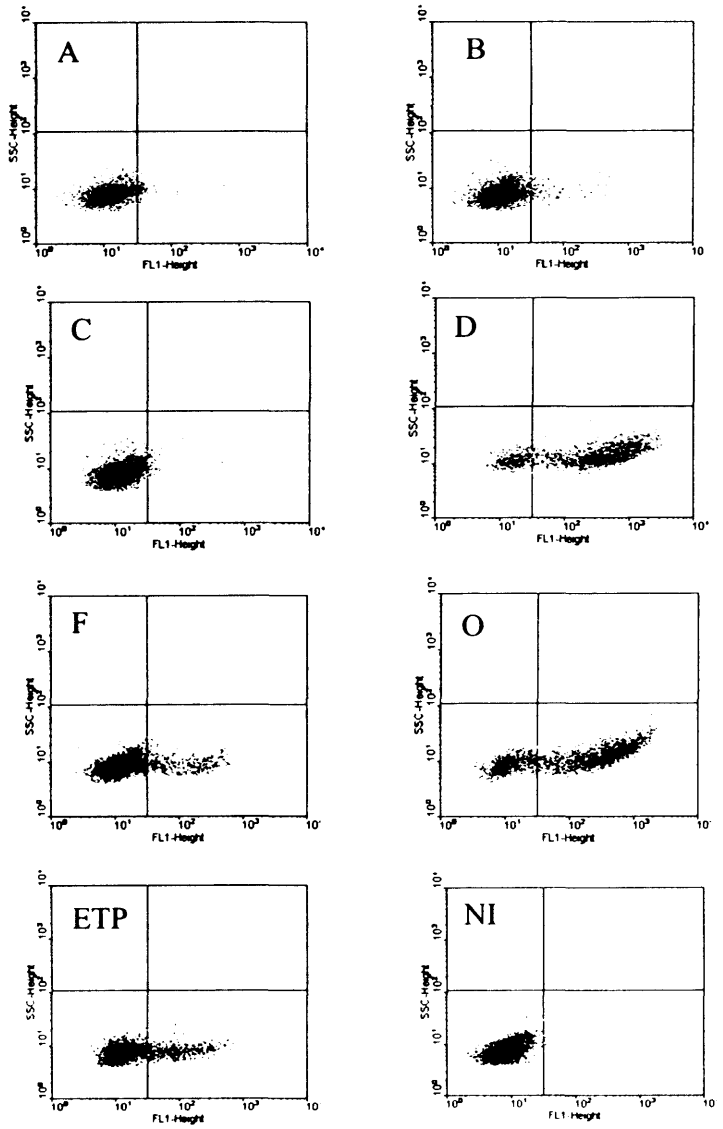


Figure 3.11 FACS plots of CVG-37 cells infected with HIV-1 subtypes A, B, C, D, F, group O, HIV-2 ETP and no infection control.
The cells in the lower right quadrant are expressing GFP and therefore infected. Uninfected cells are found in the lower felt quadrant. This data demonstrates the magnitude of GFP expression, as assessed by FACS, and that infected and uninfected cells can be clearly delineated.



infections. No HIV-2 ETP p24 data are presented, as the relevant antibody was not available.

3.3.3.3 Drug sensitivity

Given the development of the reporter cell line CVG-37 and its proven utility in studying the replication of diverse HIV-1 and HIV-2 isolates, the possibility of using this cell line as a drug sensitivity assay was investigated. This would both expand the repertoire of information gathered on the genetic variants of HIV studied, and provide a useful and more rapid method than traditional drug sensitivity assays (Gervaix *et al.*, 1997).

The antiretroviral drugs (ARVs) chosen were zidovudine (AZT), efavirenz (EFV) and ritonavir (RTV) - a nucleotide reverse transcriptase inhibitor (NRTI), a non-nucleotide reverse transcriptase inhibitor (NNRTI) and a protease inhibitor (PI), respectively. The assay was first attempted by infecting CVG-37 cells at an MOI of 0.01 with the virus NL4-3 (by spinoculation in the presence of the relevant ARV) as the infection efficiency with this virus was previously shown to be high. Titration of each drug onto CVG-37 cells, post-infection, in concentrations ranging from 0 to 0.5 μM, however, produced unusual dose response curves (Figure 3.12). Specifically, AZT and EFV inhibited replication of NL4-3 at concentrations far lower than might be expected. It was concluded that this may be a result of spinoculation in media containing drug, in effect concentrating the ARVs onto cells and enhancing any inhibitory effects expected. In addition RTV appeared to have no effect on replication of NL4-3. Experiments with RTV (and other protease inhibitors) were therefore not continued, although a two-step protocol is in development (Pirounaki *et al.*, 2000).

The protocol was modified accordingly, with an increased number of titration points in the lower concentration range (between 0 and 0.01 μM). The length of time the infected cells were left to incubate in the presence of drug was also increased (from 4 to 7 days), as it had been observed that the primary isolate viruses would replicate more slowly than NL4-3. Furthermore, the assay was only performed using HIV-1 subtypes B, D, group O and HIV-2 ETP. This is because the subtype A, C and F isolates did not achieve an appreciable level of infection (in terms of GFP expression) by 7 days PI.

Figure 3.12 Graphs showing titration of AZT, EFV and RTV on CVG-37 cells infected with NL4-3.
(A) Chart shows raw data, in terms of % cells GFP positive 4 days PI, culture in the presence of a concentration range of AZT, EFV and RTV. **(B)** Chart shows the normalised data, required for IC50 calculation. For clarity, the X axis is not extended to 0.5 µM.

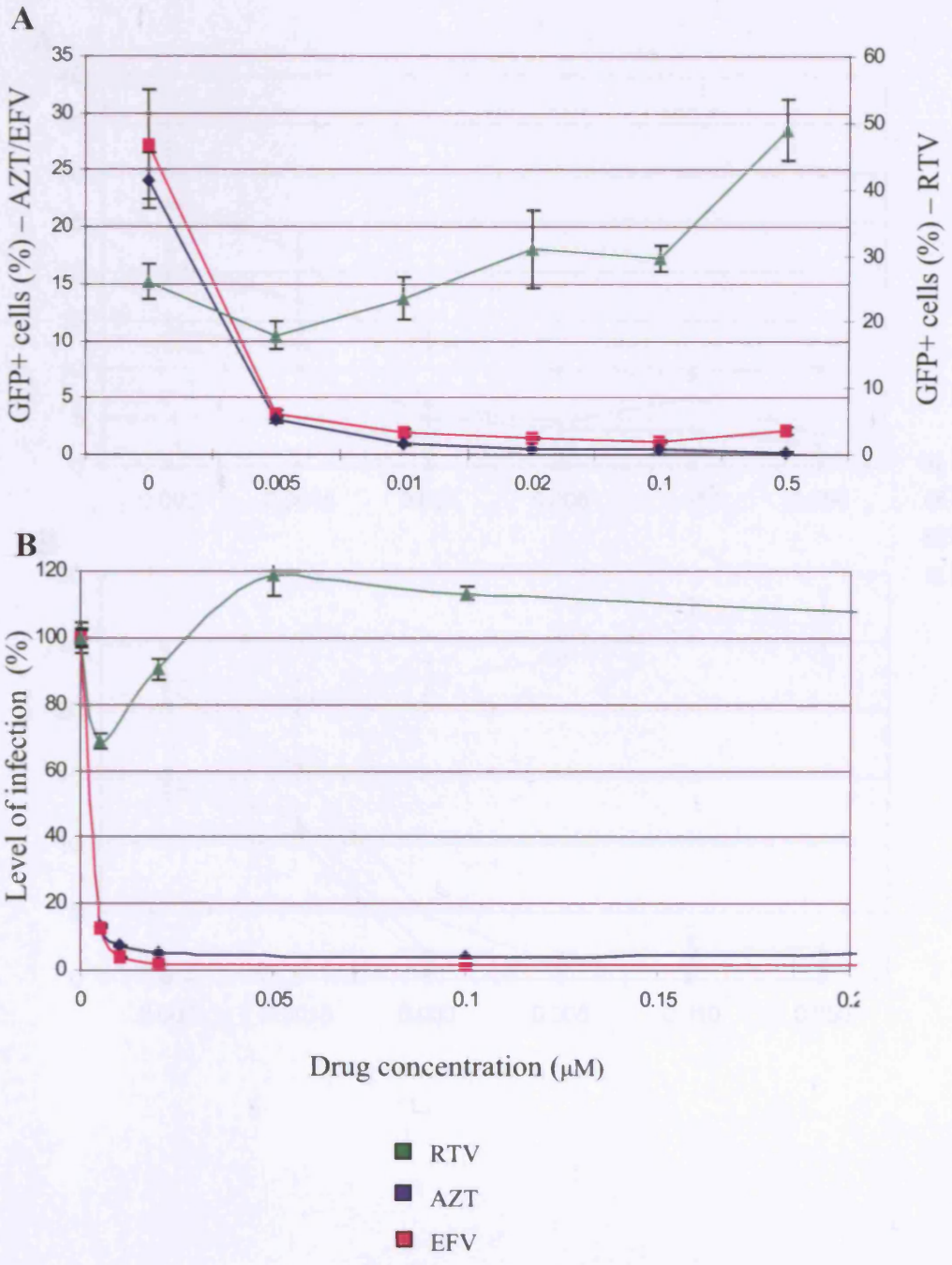


Figure 3.13 - a. Graphs showing titration of AZT and EFV on CVG-37 cells, infected with HIV-1 subtypes B, D, O and ETP (raw data).
(A-B) Charts show raw data, in terms of % cells GFP positive 7 days PI with HIV-1 subtypes B, D, group O and HIV-2 ETP, cultured in the presence of a concentration range of AZT **(A)** and EFV **(B)**.

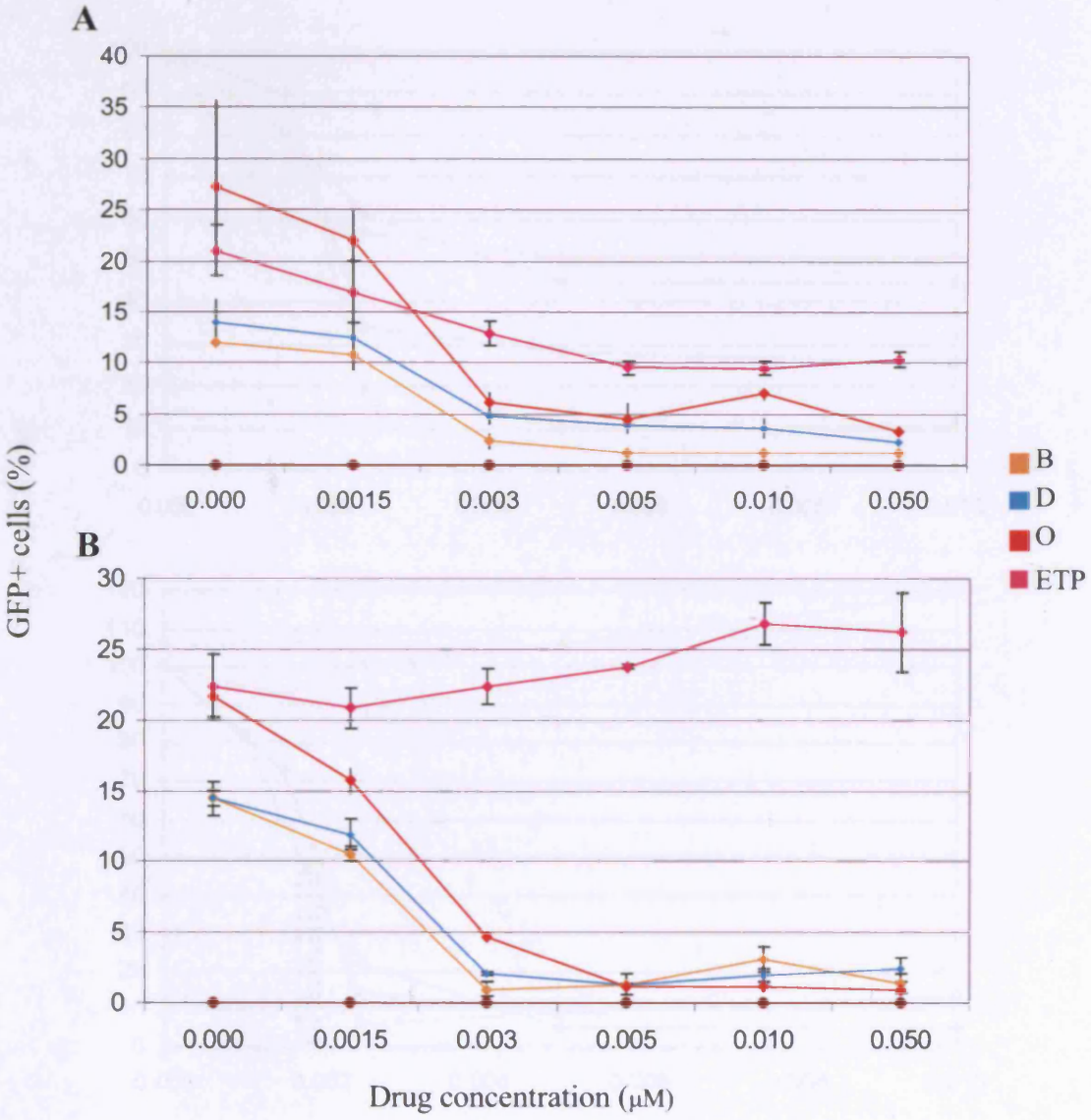
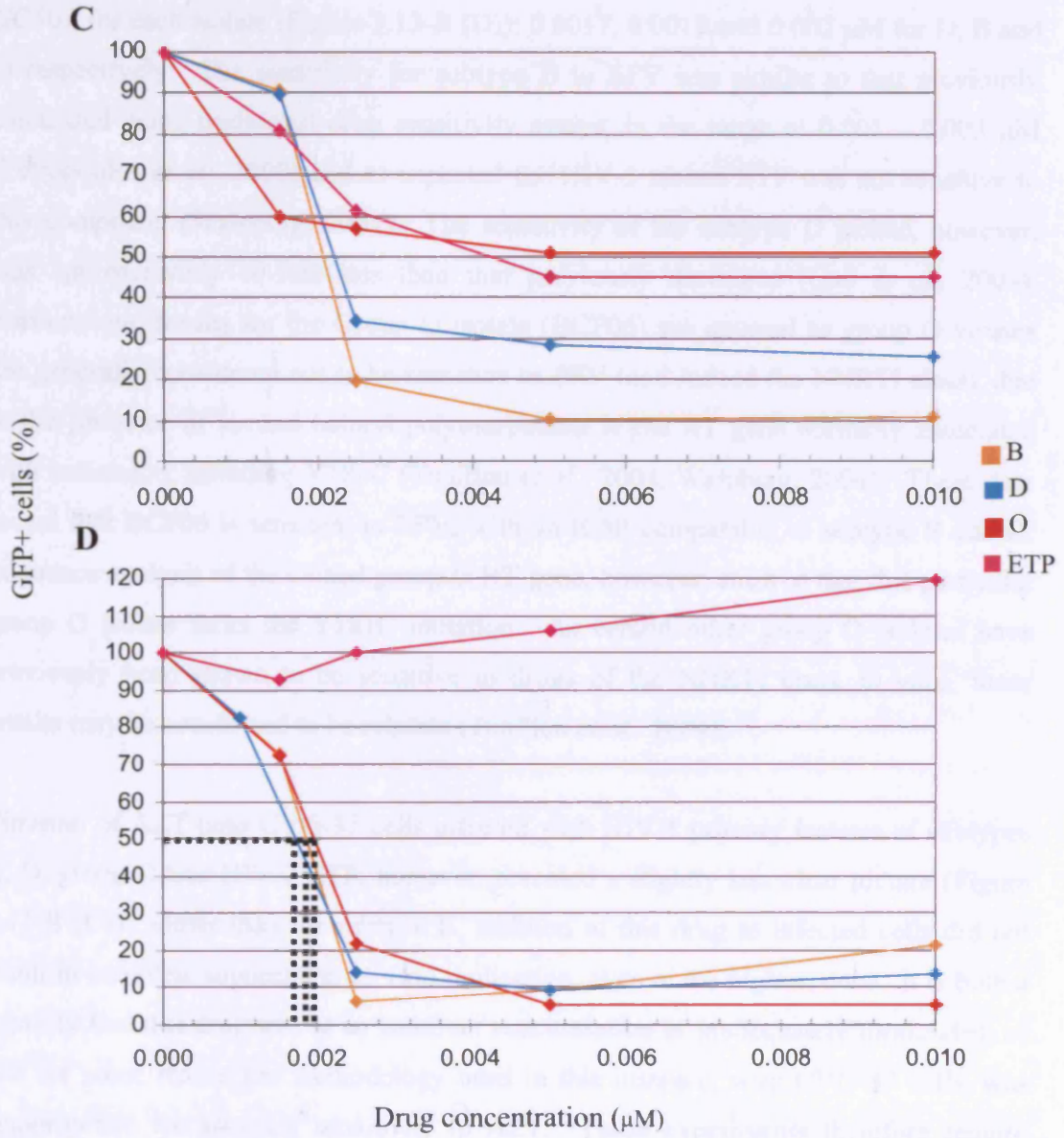


Figure 3.13 - b. Graphs showing titration of AZT and EFV on CVG-37 cells, infected with HIV-1 subtypes B, D, O and ETP (normalised data).
(C-D) These charts show the normalised data, required for calculation of the sensitivity of HIV-1 subtypes B, D, group O and HIV-2 ETP, to the ARVs AZT and EFV. For clarity, the X axis is not extended to 0.5 µM.



A new stock of HIV-1 subtype B was generated and included in these experiments.

Titration of EFV onto CVG-37 cells infected with HIV-1 primary isolates of subtypes B, D, group O and HIV-2 ETP enabled calculation of the concentration of drug at which the level of infection was reduced to 50% of that recorded in the absence of drug (IC50), for each isolate (Figure 3.13-B (D)): 0.0017, 0.0019 and 0.002 μM for D, B and O respectively. The sensitivity for subtype B to EFV was similar to that previously calculated using traditional drug sensitivity assays, in the range of 0.001 – 0.003 μM (Petropoulos *et al.*, 2000) and as expected the HIV-2 isolate ETP was not sensitive to this compound (Wainberg, 2004). The sensitivity of the subtype D isolate, however, was approximately 10-fold less than that previously described (Gao *et al.*, 2004). Furthermore, results for the Group O isolate (BCF06) are unusual as group O viruses are generally considered not to be sensitive to EFV (and indeed the NNRTI class), due to the presence of several natural polymorphisms in the RT gene normally associated with resistance, including Y181C (Tuaillon *et al.*, 2004, Wainberg, 2004). These data reveal that BCF06 is sensitive to EFV, with an IC50 comparable to subtype B and D. Sequence analysis of the cloned group O RT gene, however, showed that this particular group O isolate lacks the Y181C mutation. As certain other group O isolates have previously been shown to be sensitive to drugs of the NNRTI class, *in vitro,* these results may be considered to be reliable (Tuaillon *et al.*, 2004).

Titration of AZT onto CVG-37 cells infected with HIV-1 primary isolates of subtypes B, D, group O and HIV-2 ETP, however, revealed a slightly less clear picture (Figure 3.13-B (C)). Other than for subtype B, addition of this drug to infected cells did not result in complete suppression of viral replication, even at the highest dose. It is both a possibly that this drug was at an incorrect concentration or inadequately formulated, or that for some reason the methodology used in this instance, with CVG-37 cells, was inappropriate for assaying sensitivity to AZT. These experiments therefore require repeating, in direct comparison with a standard MT2 drug sensitivity assay, with the same viruses. This should overcome any potential problems of incorrectly formulated drug and allow optimisation of the CVG-37 assay, which is in progress.

3.4 Discussion

Differences between HIV-1 and HIV-2 in terms of disease progression and transmissibility have been clearly documented: HIV-2 exhibits a lower rate of heterosexual transmission, vertical transmission is rare and progression to AIDS is generally prolonged (De Cock *et al.,* 1993). The question remains as to whether similar differences exist among the various genetic subtypes of HIV-1. Certain studies, both *in vitro* and epidemiological, have provided indications that there may be biological differences among HIV-1 groups and subtypes, but also some apparently contradictory results (Hu *et al.,* 1999). Independent of viral subtype, rate of disease progression varies from person to person and is influenced by multiple virological and host factors. The evaluation of associations between HIV subtypes and disease phenotypes is therefore very difficult, the challenge being in determining whether any associations identified are causal, or due to bias or chance. Nonetheless, it would be premature to conclude that the genetic variability of HIV, defined as subtype, does not play any role in differences in virus biology, transmission or the development of disease (Peeters, 2001).

Whilst several different approaches have been adopted to evaluate and thus better understand the differences between the subtypes and groups of HIV-1, this study has involved the investigation of the differences between a panel of primary virus isolates, *in vitro.* By taking a simple approach and characterising the basic phenotypic differences between HIV-1 primary isolates of different subtype, in different cellular environments, the aim was to determine if there are any defining and reproducible phenotypic discriminants between examples of the HIV subtypes and whether this could be considered to agree with any of the previously published *in vitro* data. Whilst the analysis of growth phenotype cannot be directly extrapolated to the *in vivo* situation, it is reasonable to draw broad parallels between the biological phenotype of HIV isolates and potential disease phenotype, as certain characteristics are well documented independent markers of clinical outcome (Koot *et al.,* 1993, Karlsson *et al.,* 1994, Connor *et al.,* 1997). Furthermore, the comparison of HIV-1 subtypes A, B, C, D, F, group O and HIV-2 viruses in a new reporter T-cell line, CVG-37, provides a prototype system for the continuation of such studies. Given the apparent lack of consensus with

regards to the importance of HIV genetic variation on viral phenotype, this study is timely in its consideration of such matters.

3.4.1 Comparison of HIV primary isolates in vitro using p24 and microscopy

Initial growth of viruses through PMBCs in order to generate stocks, revealed that each of the viruses: HIV-1 subtypes A, B, C, D, F, group O and HIV-2, grew with particular efficiency, some considerably more rapidly than others and with distinctive CPE. As these viral stocks were obtained externally and titre was unknown, it was possible that the viral inoculum used for these preliminary infection experiments may have been sufficiently different to generate the observed differences. Calculation of virus titre for each stock prepared, however, revealed that the titres of all viruses were consistently low – in the range of 10^3-10^4 infectious units (calculated on NP2 cells) per ml. Phenotypic differences between subtypes were maintained on NP2 cells: the X4 or dual tropic genotypes (HIV-1 B, D, O and HIV-2 ETP) were fusogenic and resulted in large syncitia formation as infection progressed; the R5-tropic viruses (HIV-1 A and C) did not produce such distinct CPE. This correlates with the 'X4=SI, R5=NSI' rule, but this is only truly applicable to MT2 cells. HIV-1 subtype F, however, was different in that infection using stocks which showed a preference for R5 coreceptor usage, resulted in a heavy burden of syncitia formation during PBMC infection. Furthermore, in two of the four working stocks of HIV-1 subtype F, dual coreceptor usage was detected during titration on NP2 cells (these stocks also induced the same extensive syncitia on PBMCs and NP2 cells). This indicates that coreceptor usage within the viral population in these stocks is not particularly restricted, and that growth through PBMCs from different donors may have resulted in outgrowth of minority species that use the CXCR4 coreceptor. Cloning and sequencing of several genomic fragments from provirus within infected PBMC DNA enabled confirmation of the subtype identity of the viruses being studied.

Growth phenotypes of the different HIV-1 primary isolates were therefore further characterised in PBMCs and the T-cell line SupT1.R5 (ST1-R5), and results revealed that the differences observed during viral stock preparation and titrations were largely perpetuated. Infection was performed at a set multiplicity; infectious titre calculated as focus forming units per volume of supernatant on NP2 cells. Previous studies have used a standardised amount of p24 for equalized infections, but due to the genetic

heterogeneity of the viruses studied it was felt that this was inappropriate. Whilst RT activity of supernatant provides an alternative means to equalise viral input, it has recently been shown that the relationship between RT activity and infectivity of a viral stock, whilst well correlated for R5 tropic viruses, are not well correlated for X4-tropic viruses (Marozsan *et al.*, 2004). As viruses of both tropisms are being compared in this study, absolute infectious titre on NP2 cells expressing either the X4 or R5 coreceptors was deemed the most appropriate way of equalising infection in growth comparison experiments.

By monitoring infection of PBMCs and ST1-R5s by microscopy, it was shown that the X4- or dual-tropic viruses HIV subtype B, D and O cause a heavy burden of CPE, largely in the form of syncitia formation. HIV-1 subtypes A and C failed to induce syncitia, yet still caused pyknosis and apparent degradation of individual cells. The SI phenotype of HIV-1 subtype F was also evident in both PBMCs and the ST1-R5 cell line. Corresponding p24 data for the infections yielding these results, however, did not seem to necessarily reflect the extent of infection observed *in vitro*. For example, subtypes B, C, and F infections were clearly productive in terms of cell killing, yet the p24 analysis did not reveal as high a level of this antigen in the cell-free supernatant as one might expect. This could be due to reduced assay efficiency, the anti-p24 antibody having been generated by immunisation of mice with a subtype B epitope. Alternatively it is possible that, for different viral subtypes, the ratio of infectious to non-infectious virus particles varies. In the case of subtypes A and C, as isolates of this subtype are R5-using and have been reported to be typically 'slow' in terms of their growth *in vitro*, however, it may be that looking for p24 up to 6 days post-infection was too early when infecting with this virus at an MOI of 0.01. It was therefore concluded that, in trying to perform such comparative studies, a system was required that did not require p24 detection as its output, and preferably would allow infected cell visualisation and direct counting. This would surpass the measurement of p24 as a marker of progression of infection as, in addition to the flaws previously outlined, the p24 assay measures aggregate production and accumulation of viral proteins from cultured cells and does not permit an accurate quantitation of the actual number of infected cells.

3.4.2 Development of a reporter cell line for in vitro study of HIV primary isolates

The establishment of a CEM-GFP stable cell line has previously enabled the monitoring of X4-tropic HIV infection and antiretroviral drug sensitivity, *in vitro* (Gervaix *et al.*, 1997). This cell line and most immortalised T-lymphoblastoid cell lines, however, do not endogenously express CCR5, a member of the seven-transmembrane G-protein receptor family which is present at the surface of primary CD4+ T cells and monocytes permitting entry of primary macrophage and R5-tropic strains of HIV (Deng *et al.*, 1996, Doranz *et al.*, 1996). Accordingly, the CVG-37 cell line was developed from this reporter cell line by addition of the CCR5 coreceptor. Two clonal lines with levels of CCR5 expression equivalent to that typically detected on PBMCs were identified and their sensitivity to infection determined. This enabled selection of a clonal line with low constitutive background fluorescence, but a high level of GFP expression upon infection with HIV-1 and HIV-2 (either molecular clone or primary isolate) of either X4 or R5 tropism (Section 3.3.2.2). This cell line therefore provided a tool with which to compare the growth of the diverse range of HIV isolates under investigation in a fast, easy, inexpensive, and accurate manner. The efficiency of transactivation of the LTR-GFP construct should be comparable between different viral isolates as, despite the high level of sequence variation between HIV-1 and HIV-2, the function of Tat transactivation has been shown to be highly conserved between lentiviruses (Hooker *et al.*, 2002) and GFP expression is efficiently and strongly induced upon infection with viruses of both types. The use of these cells in drug sensitivity assays was also demonstrated, and the development of this system for more extensive studies of drug sensitivity of HIV-1 genotypic variants is underway.

Infection of CVG-37 with all isolates at a multiplicity of infection of 0.01 ffu/cell revealed that HIV-1 subtypes A, D, F, group O and HIV-2 ETP spread through the culture equally until 5 days PI, but from day 5 onwards the pattern of spread of these viruses diverged, with HIV-1 subtype D and group O continuing to spread, reflecting previous observations in the ST1-R5 cell line. HIV-2 ETP also spread but to a lesser extent, but HIV-1 subtypes A and F appeared to decline and failed to infect more than 10-15% of the CVG-37 cells. Similarly to the growth experiments on ST1-R5 cells, HIV-1 subtypes B and C failed to establish significant levels of infection - only 5% of cells expressed GFP as a result of infection, over the 12-day time course. These observations were highly reproducible. Regardless of the overall level of infection,

however, in response to all subtypes mean green fluorescence values in GFP+ infected cells were equivalent, indicating that all viral subtypes and types were capable of efficiently transactivating the LTR-GFP construct. Measurement of p24 in the supernatant at each time point was largely in agreement with the pattern of GFP data, and supports the notion that increasing GFP expression is a product of ongoing viral replication. The levels of p24 detected in this series of experiments, however, was much greater than that detected in the ST1-R5 infections, most probably attributable to the increased efficiency of initial infection using spinoculation.

3.4.3 HIV-1 primary isolates of different subtypes grow differently in vitro

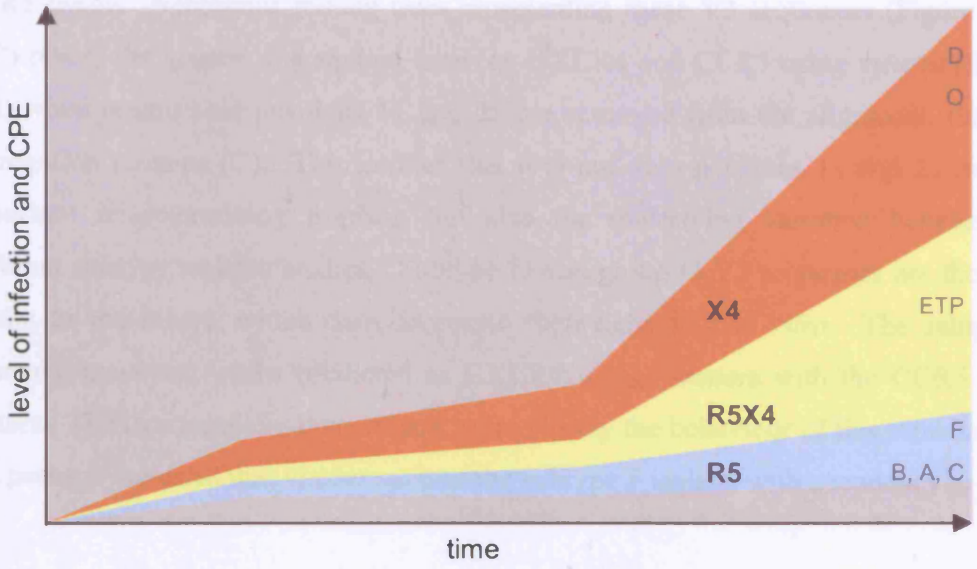
When taken together the GFP and p24 data describing the growth of HIV-1 subtype A, B, C, D, F, group O and HIV-2 ETP reveal that, in the CVG-37 cell line, there are substantial and reproducible differences in the patterns of viral replication and spread between distinct viral types and subtypes. Furthermore comparison to growth experiments performed in a different T-cell line, ST1-R5, reveals a significant correlation between the replication efficiencies of the viruses studied, particularly HIV-1 subtype D, and group O. This implies that, in the T-cell environment, the primary isolates of subtype D and group O had a greater replicative capacity than the isolates of HIV-1 subtypes A, B, C and F. The reasons for these differences, however, are by no means clear. As described in section 3.1.2, studies of the HIV LTR, *protease, gag* and accessory genes; *rev, vpu* and *tat*, have produced evidence that implies sequence variation within these genes may be manifested phenotypically, detected as changes in the way in which different viral subtypes grow *in vitro*, and thus perhaps in the way they behave *in vivo*. These effects, however, are typically subtle and whilst cumulatively they may help account for the observations made in this study, the individual contributions of subtype-specific variation within each gene cannot be determined. A considerable amount of research, however, has addressed what is a clear relationship between viral coreceptor usage, growth phenotype *in vitro* and disease progression *in vivo*, but few efforts have been made in terms of evaluating the association between subtype and coreceptor usage. The data presented here show that there is a clear delineation in growth phenotype on T-cells, between the X4 and R5X4 using isolates (subtype D, O and ETP), and R5 using isolates of subtypes A and C, subtype F being an exception (Figure 3.14). As the different viral subtypes have previously been suggested to favour use of certain coreceptors (Tscherning *et al.*, 1998)

with this experimental observation, certain explanations may be proposed as to why these viruses of different subtype grow with such reproducible differences, *in vitro*.

3.4.4 Subtype dictates coreceptor usage and therefore growth phenotype in vitro

The reasons why X4-tropic viruses grow more rapidly *in vitro* are not well understood. As gp120 sequence seems to drive this phenotype it may be hypothesised that the sequence characteristics of an *env* which confer X4-usage, may also be those which are responsible for enabling cells infected with an X4-tropic virus to have an increased capacity for HIV replication; perhaps by enhanced fusogenicity improving virus entry or release from cells. Alternatively, by interacting with the CXCR4 coreceptor the cell signalling resulting from viral engagement may be more advantageous for subsequent viral entry, replication and egress. Also, by targeting this receptor the virus may avoid certain cellular pathways that are refractory to infection. Regardless of the mechanism, however, the differences between envelope-associated viral phenotype *in vitro* would seem to be an experimental phenomenon that reflects the difference between the envelope sequences and glycosylation state. As these sequence features are highly predictive of clinical progression in a large proportion of people, it seems likely that this will hold true in the various non-B subtypes (Moore *et al.*, 2004).

Figure 3.14 **A generalised representation of the pattern of viral growth (adapted from p24 and GFP data) and extent of CPE caused by primary isolates of different HIV-1 subtypes and HIV-2 (ETP) over time, in CD4+ T-cells lines.**
The coloured sections emphasise the relationship between viral growth rate, capacity to induce syncitia formation, coreceptor usage and subtype.



Given that the genetic differences between subtypes appear to have resulted in different *env* tropism, one may postulate that this explains their differing behaviour *in vitro*, and thus potentially may result in differences between them, *in vivo*.

Whilst the mechanistic relationship between coreceptor usage and phenotype *in vitro* is not well understood, however, many of the features of the gp120 envelope viral protein involved in coreceptor usage have been revealed. The sequence of the V3 region is highly associated with the coreceptor phenotype, together with the V1V2 region (Pollakis *et al.*, 2001): the overall amino acid charge appears to dictate coreceptor usage with higher positive charges being associated with the SI phenotype and utilisation of the CXCR4 coreceptor (de Jong *et al.*, 1992a, de Jong *et al.*, 1992b), in particular at positions 11 and 25 of the V3 loop. A low frequency of V3 glycosylation has also been associated with CXCR4 usage, and the N-linked glycosylation site downstream of the first V3 cysteine has been shown to be a major determinant influencing viral replication phenotype (Pollakis *et al.*, 2000). These features thus have enabled the creation of several computational methods of predicting coreceptor usage, based on Env sequence. An alignment of V3 sequences of the viruses studied is shown in Figure 3.15-A. Using a number of methods for predicting coreceptor usage (Jensen *et al.*, 2003, Pillai *et al.*, 2003), it seems the *in vitro* observations made here reflect the *in silico* predictions, in that the subtype D and group O viruses are predicted to and do use the CXCR4 coreceptor, and are those which grew most efficiently through culture. Subtypes A and C are predicted to use CCR5 and their growth *in vitro* reflected this. Subtype F, however, is predicted as using CXCR4, whereas *in vitro* it showed a preference for CCR5 usage. Neighbour joining trees representing these V3 sequences (Figure 3.15-B/C) reveal the genetic segregation between CXCR4 and CCR5 using viruses (B) and that, when amino acid positions 11 and 25 are removed from the alignment, that this segregation remains (C). This implies that it is not only positions 11 and 25 that are important in determining tropism, but also the underlying variation between the different subtype isolates studied. Subtype D and group O V3 sequences are the most distant to the others, which correlates with their behaviour *in vitro*. The subtype F sequence however, whilst predicted as CXCR4 using, clusters with the CCR5 using viruses. The tree topology thus reflects more closely the behaviour of this virus *in vitro* and perhaps indicates that viruses such as the subtype F isolate, with a tendency to

Figure 3.15 The relationship between subtype, V3 loop genotype and phenotype.
(A) Alignment of V3 loop sequences of each of the viral isolates studies, except HIV-2 ETP, and their predicted coreceptor usage*. Position 11 and 25 are highlighted (blocks) to indicate the residues most important in the 'charge rule' of predicting coreceptor usage. Positions 24, 27 and 32 implicated in contributing to the X4 phenotype (Jensen *et al.*, 2003) are marked (red arrows). *Prediction of coreceptor usage was performed using a PSSM based method (WebPSSM, M Jensen, University of Washington), the Charge Rule, and three machine learning techniques: Support Vector Machine (SVM); C4.5; and PART, all provided via WetCat (S Pillai and B Good, University of California, San Diego). **(B)** Neighbour joining tree of V3 sequences shown in (A). **(C)** Neighbour joining tree of V3 sequences shown in (A), with positions 11 and 25 removed.

A

Virus	V3 sequence alignment	CoR*
SF162(B)	CTRPNNNT-RKSITI--GPGRAFYA-TGDI-IGDIRQAHC	R5
92BR025 (C)	CTRPNNNT-RKSIRI--GPGQAFYA-TGEI-IGDIRQAHC	R5
92UG037 (A)	CTRPNNNT-RRSVRI--GPGQTFYA-TGDI-IGDIRQAHC	R5
92TH001 (AE)	CTRPSNNT-RTSINI--GPGQVFYR-TGDI-IGDIRKAYC	R5
93BR020 (F)	CTRPNNNT-RKRISL--GPGRVFYT-TGEI-IGDIRKAHC	X4
92UG001 (D)	CTRPYYNQIRQRTXI--GQGQALYT-TRVT--GDIRKAYC	X4
BCF06 (O)	CMRKGRGK-IQRIAT-GPLRWVSMAAKTESQNTGSRIAYC	X4

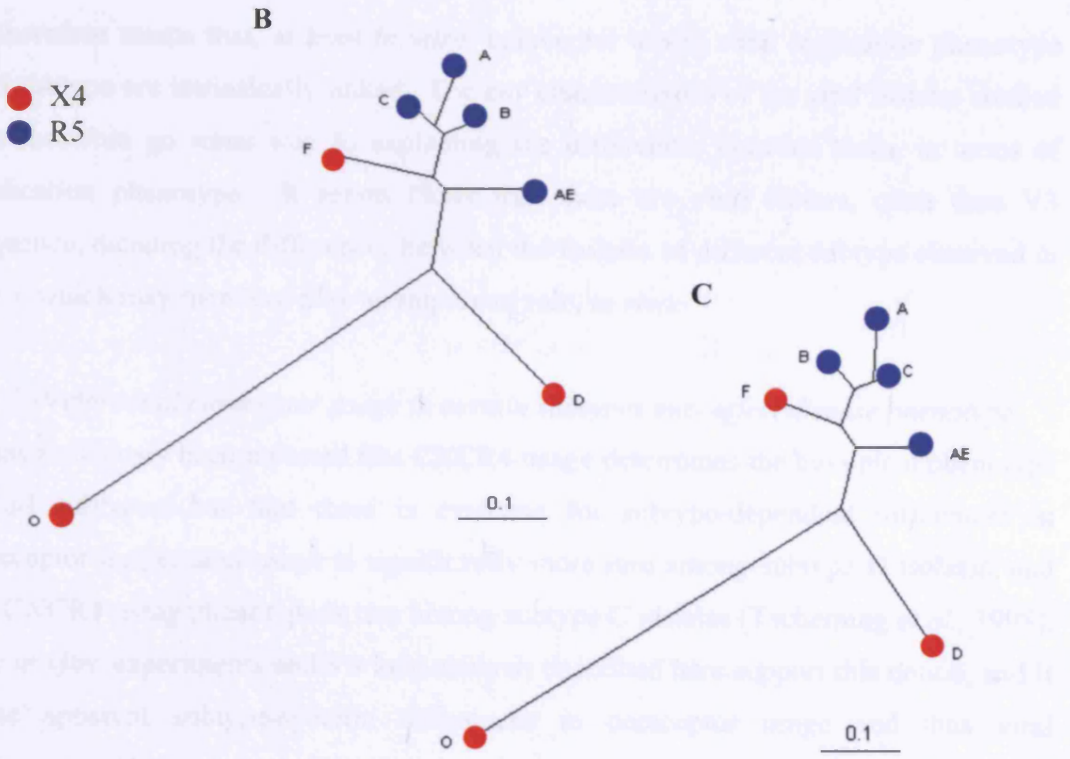


exhibit dual tropism, use a different 'sequence key' to gain access to T-cells via CXCR4, as they have retained that ability to use CCR5. Accordingly, in these cases the standard 'rules' may not apply in coreceptor usage prediction (Pillai *et al.,* 2003).

Alternatively, these data could indicate that the subtype F virus, rather than using a different sequence 'key', is merely further along the pathway of accumulated mutations that result in CXCR4 usage (Jensen and van 't Wout, 2003). PSSM analysis (Jensen *et al.,* 2003) allows the calculation of a V3 'score' which, rather than being a binary definition, places viruses on a continuous scale from R5- to X4-using and accordingly takes into consideration the background variation in V3 sequence outlined in this small study as being important. Analysis of the subtype F sequence gives a score of –5.75, which compared to the scores for R5-using subtype A and C (-7.49 and –10.67 respectively) and X4-using subtype D (8.33) perhaps places this virus marginally closer to X4-usage, in terms of score. It should be noted, however, that this tool is designed for analysis of subtype B sequences.

It therefore seems that, at least *in vitro,* coreceptor usage, viral replication phenotype and subtype are intrinsically linked. The *env* characteristics of the viral isolates studied can therefore go some way to explaining the differences between them, in terms of replication phenotype. It seems likely that there are viral factors, other than V3 sequence, dictating the differences between the isolates of different subtype observed *in vitro,* which may therefore play an important role, *in vivo.*

3.4.5 Preferential coreceptor usage in certain subtypes may affect disease phenotype

It has previously been reported that CXCR4 usage determines the biological phenotype of all subtypes, but that there is evidence for subtype-dependent differences in coreceptor usage: dual usage is significantly more rare among subtype D isolates, and the CXCR4-using phenotype is rare among subtype C isolates (Tscherning *et al.,* 1998). The *in vitro* experiments and V3 loop analysis described here support this notion, and if these apparent subtype-specific differences in coreceptor usage and thus viral replication capacity are reflected *in vivo*, this creates the possibility that genetic subtypes may differ in clinically important properties such as virulence, tissue tropism and transmissibility.

The most striking conclusions that may be drawn from the review of literature in sections 3.1.1 – 3.1.2 are that: subtype D viruses have been associated with higher viral loads, rate of disease progression, in some cases transmission frequency, and typically grow well in culture (in primary and T-cell lines) with a preference for CXCR4 usage; and that subtype C viruses have been identified as particularly slow growing in culture, with a preference for R5 usage, yet have received particular attention in terms of research as they currently dominate the HIV-1 pandemic. Taking these observations into consideration, it would seem that the *in vitro* data presented here agree with these data and it is tempting to hypothesise that in the case of subtype D, growth phenotype *in vitro* does in fact reflect that manifested *in vivo*. For subtype C, however, it seems that there is a dichotomy between epidemiological observations and *in vitro* findings, which is most likely to be due to the growth disadvantage that R5-tropic viruses seem to suffer in T-cells, *in vitro*. It is therefore important to note that no presumption is made in this study that the attenuated replication capacity of R5-tropic viruses compared to X4 viruses, *in vitro,* reflects a real life attenuation of these isolates. In addition to the potential explanations made in section 3.4.4, it may be genetic variants of HIV which are characterised as CCR5 using require additional factors for productive infection in culture not provided in experimental systems used here, such as additional coreceptors or a particular form of cell-cell contact in a certain cellular environment. X4-tropic viruses may not require this. It may also be that the T-cell lines used in this study express high levels of β-chemokines (the natural ligands for CCR5) either endogenously or in response to infection, which impede the progression of R5-using viruses by reducing coreceptor availability. Thus, data generated in the experimental system used in this study must not be over interpreted. It is tempting to suggest, however, that the subtype D data are intriguing and it will be interesting to see if, in the future when group O viruses have been further studied, whether they too show enhanced disease progression *in vivo,* or like HIV-2, which in CVG-37 cells grew better than several of the HIV-1 subtypes but *in vivo* has a much attenuated clinical course, the results obtained in this study are merely a reflection of this genotype's ability to gain access to T-cell by an as yet undetermined mechanism.

In vivo, the presence of X4-tropic virus in HIV+ patients has long been observed as indicative of a poor prognosis (Cheng-Mayer *et al.*, 1988, Fenyo *et al.*, 1988, Tersmette *et al.*, 1988), as whilst approximately 50% of isolates from patients with AIDS have an X4-tropic phenotype, identification of X4-tropic viruses at seroconversion is relatively rare. Whether this is a feature of a selective bottle neck at transmission resulting in positive selection of R5-tropic viruses, and an indication that X4 tropic viruses herald a decline in CD4+ count associated with the onset of AIDS because they are inherently more pathogenic is unclear. At present, it seems that the common perception is more along the lines that chronic immune activation is the proximal cause of T-cell loss, which facilitates the outgrowth of X4 tropic viruses in the later stages of disease when sites of the body (for example gut associated lymphoid tissue - GALT) where R5-tropic viruses had more successfully colonised earlier in infection, become exhausted (Grossman and Paul, 2000, Moore *et al.*, 2004). The question remains, however, if the viral subtype which you harbour is closer in terms of both evolutionary distance and time to switching to CXCR4 usage, is your prognosis worse? The potential prognostic value of 'V3 scoring' has been described previously (Jensen and van 't Wout, 2003). It may be, therefore, that if further studies were performed which confirm previous observations and bear out those made in this study with regards to the relationship between viral subtype, growth *in vitro* and coreceptor usage, then viral subtype may also be of prognostic value.

Chapter 4.0 Transcriptional profiling of HIV infection and the significance of genotypic variation

4.1 Introduction

Genomics represents the study of whole sets of genes or gene products, rather than individual genes. With the completion of the draft human genome sequence (Venter *et al.*, 2001, International Human Genome Sequencing Consortium, 2004) and the availability of whole genomes of over 1000 viruses and 100 microbes (Entrez Genome, August 2004), many of which are known human pathogens, this field has rapidly become integral to the understanding of the process of infection at the molecular level. To achieve this several novel experimental approaches and technologies have been developed, collectively termed functional genomics. DNA microarray hybridisation analysis has become one such established methodology for measuring the expression levels of whole-genome encoded mRNAs due to its simplicity, comprehensiveness, data consistency, and high throughput.

4.1.1 Microarray technology

The underlying principle of all array experiments is that labelled nucleic acid molecules in solution hybridise, with high sensitivity and specificity, to complementary sequences immobilised on a solid substrate. This enables parallel quantitative measurement of many different sequences in a mixed population (Southern *et al.*, 1999, Brown and Botstein, 1999). This hybridisation property of DNA was first exploited in 1975 when Ed Southern pioneered the technique of detecting specific DNA fragments in a mixture separated by gel electrophoresis (Southern, 1975). Widespread use of this methodology revolutionised the field of molecular biology and soon variations on the Southern theme, including Northern and Western blotting (RNA and proteins respectively) together with scaling up of the number of DNA fragments studied at one time (Lennon and Lehrach, 1991), laid the foundations for microarrays. With increasing availability of the necessary hardware, namely robots and desktop computers, microarray technology has become a widely exploited tool in the post-genomic era.

Several methods for constructing microarrays have been developed (Ramsay, 1998, Watson *et al.*, 1998), two of which predominate. In one, DNA microarrays are

constructed by physically attaching DNA fragments such as library clones, polymerase chain reaction (PCR) products or oligonucleotides to a solid substrate (Schena *et al.*, 1995). For the second, arrays are constructed by synthesising single-stranded oligonucleotides *in situ* by use of photolithographic techniques (Lockhart *et al.*, 1996). Gene expression array experiments can also be performed by hybridising a single labelled mRNA sample to "macroarrays" of DNA elements on positively charged nylon or nitrocellulose filters (Tau *et al.*, 1999, Richmond *et al.*, 1999, Cohen *et al.*, 2000a, Eckmann *et al.*, 2000, Rosenberger *et al.*, 2000). Because this format does not require any special arraying or scanning equipment, arrays can be made and analysed relatively cheaply. The major disadvantages of this format, however, are reduced sensitivity (Cohen *et al.*, 2000a), limited elements, and the need for higher concentrations of labelled cDNA. For a description of different array methodologies, see Table 4.0.

4.1.2 Microarray methodology

For DNA microarrays, relative transcript abundance can be measured in two ways, either by labelling two samples with different fluorescent dyes, hybridising them simultaneously and determining the fluorescence ratio for each spot on the array, or by hybridising a single labelled sample. In the latter case, as for oligonucleotide arrays, multiple probes from the same gene, each with a corresponding mismatch probe that serves as internal control, as well as labelled transcript of known amounts for standard genes makes quantitative measurement of transcript abundance possible after hybridising a single labelled sample (Lockhart *et al.*, 1996). For both techniques use of fluorescent labelling enhances sensitivity and the dynamic range of measurement, although radiolabelling provides the best dynamic range.

Table 4.0 **Summary of various array technologies**

	Filter Array	Microarray	Affymetrix	Rosett
Solid Support	Nylon membrane	Glass slide	Glass	Glass
Probe	PCR product or oligonucleotide	PCR product or oligonucleotide	Oligonucleotide	Oligon
Synthesis	Robotic deposition	Robotic deposition	*In situ* by photolithography	*In situ* deposit
Label	33P	Cy3 and Cy5	Fluorescein	Cy3 an
Labelling method	Reverse transcription or end-labelling	Reverse transcription or amino-allyl ligation	Reverse transcription or biotinylation	Revers allyl lig
Scanning	Phosphor imager	Fluorescent scanner	Affymetrix scanner	Fluores

Messenger RNA from eukaryotic cells is usually extracted by affinity purification of mRNA with an oligo-dT resin, followed by incorporation of dye-labelled nucleotides into cDNA molecules during reverse transcription. Typically either poly(dT), which primes from the poly(A)-tail of mature mRNA, random primers, or sequence specific oligonucleotides, either those used to construct the microarray or a minimally complex mixture of octamers sufficient to hybridise to the 3' end of every ORF (Talaat *et al.*, 2000) are used in this step. The fluorophores most commonly used are Cy3 and Cy5, but this does differ between array types. Membrane arrays employ the use of radiolabels such as ^{33}P, for example. In the case of cDNA or oligonucleotide arrays, fluorescent labels are detected either using charge coupled device (CCD) camera or photomultiplier tube (PMT) scanner. A schematic of spotted cDNA microarray methodology is shown in Figure 4.0-A.

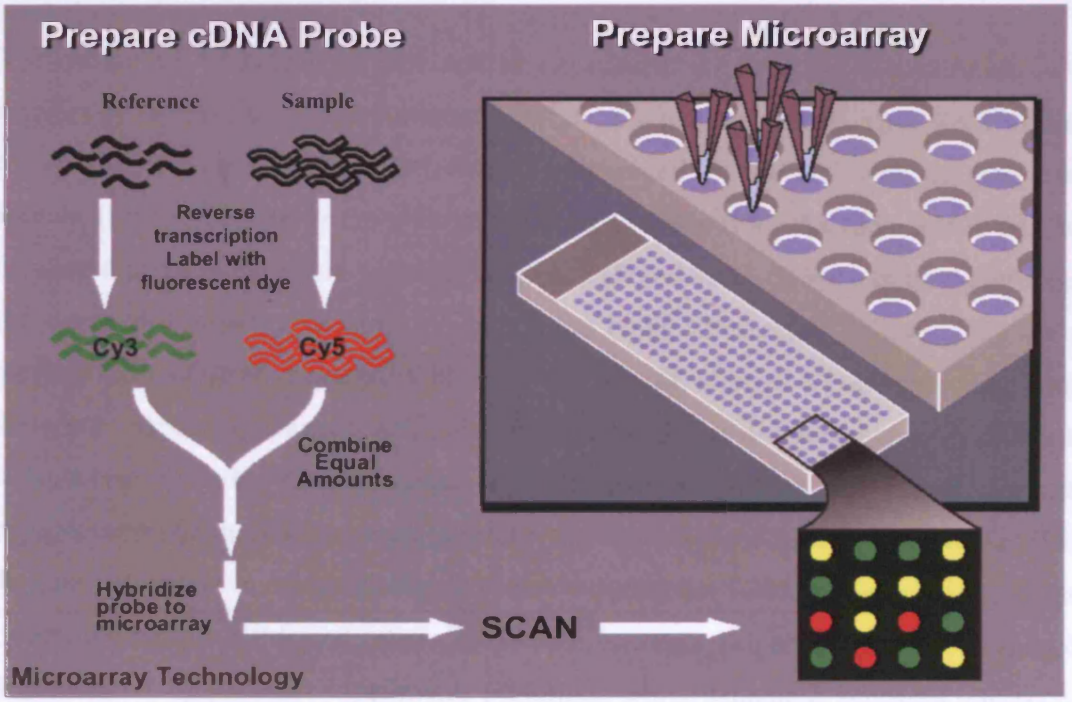
A significant limitation of current microarray technologies, however, is the large amount of input mRNA required to generate labelled cDNA for array analysis. With very small tissue samples one can either increase cDNA labelling and hybridisation efficiency, or include an amplification step to generate sufficient RNA/cDNA for use with a standard labelling method. The Affymetrix system uses a single round of amplification as a routine part of the labelling protocol. RNA amplification techniques, however, must preserve the relative abundances of the different RNAs within the starting population and be highly reproducible.

4.1.3 cDNA Microarray data analysis

The most basic cDNA microarray experiment involves the comparison of two samples, for example infected and uninfected cells. In this situation the two samples are labelled with different fluorophores, hybridised on the same microarray and the image scanned to generate an image of the array from which numerical data for each element, or spot, can be extracted. Typically these data equate to the intensity of each spot compared to the background intensity of the slide, resulting in an expression ratio for each gene, subsequently converted to log base 2 (log(2)). This transformation provides an intuitive view of the data as up- and down-regulation are of the same magnitude, and creates a normal distribution of the expression ratios (Figure 4.0-B). By depicting these data graphically as a scatterpot, with the log(2) ratio of each spot (M) against log(2) overall intensity (A), the resulting M vs. A plot enables identification of differentially

Figure 4.0 DNA microarray technology.
(A) Schematic of microarray methodology, as applied to the study of gene expression in the context of viral infection. The cDNA probe is prepared by labelling the sample (virally infected cells) and reference (normal cells) RNA with different fluorescent dyes: red for sample (Cy5) and green for reference (Cy3). These are combined in equal amounts and hybridised to a glass slide that has spotted onto it multiple cDNAs, each representing a different gene. After hybridisation the array is scanned – the composite image represents the relative abundance of mRNA for each gene on the array, in the samples. That is, if a spot is yellow, the mRNA in question is equally abundant in both conditions. If it is red, it is more abundant in infected cells and if it is green, the mRNA is more highly expressed in uninfected cells. In the case of this example, such an experiment would help determine which genes are differentially expressed in cells during viral infection. Reproduced from Darryl Leja, http://www.accessexcellence.org/AB/GG/microArray.html. **(B)** Representation of log2 transformed gene expression measurements and their derivation from raw data. Reproduced from Kellam and Liu, 2002.

A



B

Expression values							
Cy5	5000	10000	6000	20000	15000	10000	16000
Cy3	40000	40000	12000	20000	7500	2500	2000
Ratio	1/8	1/4	½	1	2/1	4/1	8/1
Log(2) ratio	-3	-2	-1	0	1	2	3

expressed genes as outliers. Multiple repetitions of this experimental format increase confidence in the data, reducing the variance (Yang and Speed, 2002) and allowing the statistical significance of observations to be established. This can be achieved by several means, including t-Tests (Wolfinger *et al.*, 2001) Z-scores (Thomas *et al.*, 2001), analysis of variance (Kerr *et al.*, 2000) and Wilcoxon rank sum (Mann-Whitney U) statistics (Chambers *et al.*, 1999). Three replicates of each data point are widely considered as sufficient (and practical) for statistical analysis, although larger numbers are always desirable (Lee *et al.*, 2000). Dye-swap experiments are also useful to control for differential incorporation of the two dyes (Tseng *et al.*, 2001), these only being relevant in the context of the specific experiment being performed.

The design of a microarray time course experiment requires the consideration of a number of factors. An important consideration is what kind of reference RNA to use: all samples subsequent to time zero may be compared (hybridised) with the time zero sample (Iyer *et al.*, 1999) or, alternatively, all samples from the time course can be compared to an independent common reference (Perou *et al.*, 1999). In the latter case, the expression ratios generated for each gene (sample/reference or Cy5/Cy3) are a measurement of gene expression level in the sample relative to the reference. This therefore enables comparison of experiments performed at different times, in different laboratories, provided that the same reference RNA is used. In the former case the expression ratios derived are more intuitive as they relate the gene expression level in the sample relative to a biologically informative reference, for example uninfected cells (where the sample is infected cells). These data, however, can never be cross-compared to data derived from other experiments performed with a separately prepared reference. Both reference options have advantages and disadvantages, but the latter method seems to be gaining favour with increasing attempts to integrate microarray data from disparate sources (Brazma, 2001).

4.1.3.1 Data normalisation

Before array data can be analysed, however, it must be normalised (median centring). This adjusts for systematic differences in the labelling and detection efficiencies (for the different labels) and differences in the quantity of RNA, and allows the comparison of gene expression levels across multiple slides. The effect of the reference on log(2) ratio values is that it shifts all values by a constant. This therefore allows for removal of the

reference effect, making gene expression ratios independent of the reference sample. This can be achieved for genes, by subtracting the median gene expression value for each row from all the values in the row, thereby making the median value for the row 0. This can be performed iteratively for every row in the gene expression matrix.

For normalisation of arrays, the most common method is to apply a scaling factor that sets the average or median expression ratio to 1 (corresponding to a log(2) ratio of 0), given a normal distribution. This 'global normalisation' assumes that the Cy3 and Cy5 intensities on a slide are related by a constant factor, and that the number of genes up-regulated balances the genes down-regulated so, on average, there is no change in median or mean gene expression levels.

Following this pre-processing of experimental data, further analysis may be performed in order to extract biological information. Many other normalisation procedures exist, for a concise review on this subject and other fundamentals of cDNA microarray data analysis see Leung and Cavalieri, 2003.

4.1.3.2 Clustering

A microarray experiment produces an enormous amount of data about genes that are up and down regulated at a given time and/or in a given condition, which is of very little use in the absence of some way to present the data in an interpretable and biologically meaningful manner. Sophisticated data exploration techniques have been developed which rely upon the assumption that shared gene expression often implies shared function (Bassett *et al.*, 1999, Eisen *et al.*, 1998, Yeung *et al.*, 2004). That is, within a dataset clustering genes into groups based on a measure of similarity in their gene expression over all samples provides a means for data exploration and visualisation.

The aim of clustering is to organise a collection of patterns into clusters, based on similarity, so that the patterns within the same cluster are more similar to each other than they are to a pattern belonging to a different cluster. This is not dissimilar to the way in which HIV PR-RT sequences were partitioned into subtype groups (Chapter 2) for example using the NJ method. The Euclidean distance is one of the more commonly used measures of dissimilarity, others including Pearson's correlation or Spearman's Rank correlation, Mahalanobis distance, Chebyshev distance and Canberra metric.

Once the pattern proximity has been calculated (based on distance), the clustering (or pattern-grouping) methods are used to construct tree structures of the data. These may be hierarchical or non-hierarchical, and unsupervised or supervised (Alon *et al.*, 1999, Tamayo *et al.*, 1999, Eisen *et al.*, 1998, Ben-Dor *et al.*, 1999). Unsupervised clustering is often followed by a form of supervised clustering where, if a co-regulated class of genes is known, supervised clustering algorithms (trained to recognise known members of the class) can assign uncharacterised genes to that class.

The clustering method used for analysis of microarray data presented here is average-linkage hierarchical clustering (Hartigan, 1975, Eisen *et al.*, 1998). This groups genes by their expression pattern across the samples (gene vector), and groups samples by the expression pattern of the genes within each sample (sample vector). Data are assembled into a matrix of log(2) expression ratios, each row representing a single gene and each column a single array. Gene and array expression vectors are compared in pair-wise fashion by their Pearson correlation coefficient. The two most similar vectors are joined by a node and the average vector calculated. This average vector replaces the two joined vectors and the process is repeated. This agglomeratively builds up a dendrogram relating all the genes or samples by their expression pattern. The shorter the branch connecting two genes or samples, the more similar the expression patterns. The tree structures force the genes and arrays in the expression matrix to be reordered so that genes or samples with similar expression patterns become clustered together. Expression levels are represented as a colour scale from green, representing negative, through black to red, red representing positive log(2) expression values which are derived directly from the initial data matrix. This provides a simple visual means for identifying groups of co-expressed genes. A means of ordering the data termed 1-dimensional self-organising maps (SOM) may then be used to order the clusters along the axis of the dendogram according to their similarity (Eisen *et al.*, 1998, Tamayo *et al.*, 1999).

4.1.4 Gene expression profiling of the host and pathogen interaction

Global profiling of gene expression is an ideal approach for assessing gene function as it is generally accepted that a gene is usually transcribed only when and where the function of its product is required. In the context of infection, by examining the locations and conditions under which a gene is expressed inferences may be made about

its role in the interaction between host and pathogen. Since high-density DNA microarrays were first described in 1995, the use of this technology has made a marked impact on a number of fields of research, including cellular physiology, cancer biology and pharmacology. Given the number of microbial genomes for which all sequence is known it is practically reasonable to also produce arrays for pathogens. Arrays for *Mycobacterium tuberculosis* (Wilson *et al.*, 1999), *Plasmodium falciparum* (Hayward *et al.*, 2000, Rathod *et al.*, 2002), *Streptococcus pneumoniae* (de Saizieu *et al.*, 2000), *Salmonella typhimurium* (Eckmann *et al.*, 2000, Rosenberger *et al.*, 2000) and *Listeria monocytogenes* (Cohen *et al.*, 2000c) have indeed been generated and used to analyse pathogen interactions with the host and response to antimicrobials. Such arrays have also proved key in annotating the genomes of these organisms and close relatives (Koonin *et al.*, 1997, Tatusov *et al.*, 2000), for the identification of candidate virulence factors (Cotter and Miller, 1998), and in microbial genotyping (Kato-Maeda *et al.*, 2001).

Microarrays have also been used for the study of viral and host gene expression, during infection (see Table 4.1 for full list of virus array studies). Some viruses, such as HCMV, HSV-1, KSHV, HPV-31, rhesus rotavirus, RSV and influenza virus, have pleiotropic effects on host gene expression, modulating the transcription of many genes across numerous functional groups. Some of the differences between expression patterns generated by different viruses could be due to cell-type specific expression patterns or differences in the array methodologies used. Despite these limitations, however, comparisons can be made between array results from cellular infection by different viruses and common expression changes can be found in host genes of certain functional groups, including: the interferon response; cytokines; stress responses; protein synthesis and the cell cycle.

Many microarray studies have revealed that infection of human cells with a diverse range of viruses leads to an increase in the expression of interferon-stimulated genes, induction of which is indicative of the activation of the cellular anti-viral response (reviewed in Goodbourn *et al.*, 2000). The function of interferon responsive genes in the anti-viral response is supported by array analysis showing that the induction of interferon-responsive genes *in vivo* in the liver of a chimpanzee during infection with HCV correlates with the clearance of viraemia (Bigger *et al.*, 2001). Some viruses,

however, are able to suppress the transcriptional induction of host anti-viral genes, for example whilst UV-inactivated HSV-1 leads to up regulation of interferon-stimulated genes, transcriptionally active HSV-1 is able to counteract this and keep the level of these genes constant (Mossman *et al.,* 2001). Coincident with the up regulation of interferon responsive genes during viral infection, increased expression of various cytokine genes is observed, the expression of certain cytokines probably being a common feature of infection by all viruses. For example IL-6 is up regulated upon infection by HSV-1, HCMV, HHV-6, HPV, influenza and coxsackievirus B3, and expression of RANTES is induced by HCMV, rhesus rotavirus, HPV-16 E6 and E7, RSV, PVM and HTLV-1 (for all references see Table 4.1). Adenoviruses, however, down regulate certain cytokines including CLL2, CXCL1, and IL-6.

Protein synthesis and the cell cycle are both affected, in different ways, by different viruses. Array analyses have shown that infection by HSV-1, reovirus serotype 3, echovirus 1 and HIV-1 lead to cell cycle arrest at the G2 to M transition by up regulation of GADD45, a cdc2 kinase inhibitor, suggesting this may be a common mechanism for virus induction of cell cycle arrest. HPV-31 infection, however, stimulates entry into the cell cycle and HSV-1 US11 inhibition of cell growth arrest and the HIPK2/p53 pathway has recently been described (Giraud *et al.,* 2004). The alternative effects of different viruses on cell cycle gene expression are therefore likely to reflect variation in viral replication strategies. The induction of stress-response genes is also common in response to infection by many viruses. Typically these genes are associated with heat shock and oxidative stress (Palmiter, 1998), often involving modulation of GADD45.

For HIV infection, DNA microarrays have been used to investigate both whole virus and virus protein-specific effects on cellular gene expression. Despite the apparent simplicity of this virus, however, many facets of its interaction with the host remain to be understood.

Table 4.1 Viruses whose effects on host gene expression have been measured using DNA arrays.

The array type used and approximate number of genes analysed is indicated (in brackets). A selection of the literature, relevant to the text, is presented.

Order	Family	Species	Array type	Reference(s)
dsDNA virus	Adenovirus	Adenovirus types 5 and 12	Membrane (18376)	Vertegaal et al., 2000
		Adenovirus type 2	Microarray (12309)	Zhao et al., 2003
	Herpesvirus	HSV-1	Membrane (588)	Khodarev et al., 1999
			Microarray (57)	Stingley et al., 2000
			Membrane (588)	Hobbs and DeLuca, 1999
			Microarray (19000)	Mossman et al., 2001
			Membrane (149)	Hill et al., 2001
			Membrane (18378)	Tsavachidou et al., 2001
			Affymetrix (4626)	Ray and Enquist, 2004
		Human cytomegalovirus (HCMV)	Affymetrix (4x1650)	Zhu et al., 1998
			Affymetrix (12626)	Browne et al., 2001
			Microarray* (8942)	Simmen et al., 2001
		HHV-6	Membrane (1152)	Mayne et al., 2001
		EBV	Membrane (4146)	Carter et al., 2002
		KSHV	Membrane (10000)	Mikovits et al., 2001
			Microarray* (4165)	Moses et al., 2002
			Membrane (2350)	Poole et al., 2002
			Microarray (9180)	
			Affymetrix (	
		Pseudorabies virus (PRV)	Affymetrix (4626)	Ray and Enquist, 2004
		Marek's disease virus (MDV)	Membrane (1126)	Morgan et al., 2001
	Papillomavirus	HPV-31	Microarray (7075)	Chang and Laimins, 2000
	Poxviruses	Modified vaccinia Ankara (MVA)	Microarray (15000)	Guerra et al., 2004
		Varicella-zoster virus (VZV),	Membrane (147)	Cohrs et al., 2003
dsRNA virus	Reovirus	Reovirus serotypes 1 and 3	Affymetrix (12626)	Poggioli et al., 2002
		Rhesus rotavirus	Microarray (38432)	Cuadras et al., 2002
ssRNA- virus	Paramyoxvirus	Measles virus	Membrane (3x1176)	Bolt et al., 2002
		Respiratory syncytial virus (RSV)	Membrane (268)	Zhang et al., 2001b
			Affymetrix (12626)	Tian et al., 2002
		Pneumonia virus of mice (PVM)	Affymetrix (12626)	Domachowske et al., 2002
	Orthomyoxovirus	Influenza virus	Microarray (4608)	Geiss et al., 2001
ssRNA+ virus	Picornavirus	Poliovirus	Microarray (10000)	Johannes et al., 1999
		Echovirus 1	Membrane (588)	Pietiainen et al., 2000
		Coxsackievirus B3	Microarray (7000)	Taylor et al., 2000
		Theiler's murine encephalomyelitis virus (TMEV)	Membrane (15?)	Palma and Kim, 2004
	Flavivirus	Hepatitis C virus (HCV)	Affymetrix (7000)	Bigger et al., 2001
			Microarray (1080)	Honda et al., 2001
			Microarray (23040)	Okabe et al., 2001
			Affymetrix (6000)	Iizuka et al., 2002
		Dengue virus	Affymetrix (22,283)	Warke et al., 2003
	Bunyavirus	Sin Nombre virus (SNV) and non-pathogenic Prospect Hill virus (PHV)	Affymetrix (12000)	Khaiboullina et al., 2004
dsDNA-RT virus	Hepadnavirus	Hepatitis B virus (HBV)	Microarray (1080)	Honda et al., 2001
			Microarray (23040)	Okabe et al., 2001
			Microarray (2208)	Wu et al., 2001a
			Membrane (12393)	Xu et al., 2001
			Affymetrix (6000)	Iizuka et al., 2002
ssRNA-RT virus	Retrovirus	HIV-1	Microarray (1506)	Geiss et al., 2000
			Affymetrix (6800)	Corbeil et al., 2001
			Membrance (1176)	Gibellini et al., 2002
		Human T-cell leukaemia virus-1 (HTLV-1)	Membrane (10000)	Mikovits et al., 2001
			Membrane (588)	de La Fuente et al., 2000
		Human foamy virus (HFV or SFVcpz(hu))	Membrane (588)	Wagner et al., 2000

* denotes the use of a microarray with only one fluorophore.

4.1.4.1 HIV and Microarrays

Several different approaches have been used in the past to study the effects of HIV on host cells, during infection. As HIV possesses few protein encoding genes, it must interact with multiple host proteins and pathways in order to facilitate it's life cycle and whilst reliance on the host for processes such as transcription and translation of the provirus are particularly obvious, other virus-host interactions taking place during infection are not so intuitive. By studying host proteins incorporated into virions initial conclusions were made (Ott, 1997), for example the presence of HLA-antigens, adhesion molecules and other cell surface proteins integrated into the virion surface enabled a link to be forged between viral acquisition of host factors and immune modulation and evasion. Studying factors within HIV particles also provided clues to the involvement of host factors including cyclophilin (Franke, 1994, Thali *et al.,* 1994, Ott *et al.,* 1995) and ubiquitin (Ott *et al.,* 1996) in the HIV life cycle, observations which prefaced some of the most recent advances in our understanding of the biology of retroviruses and their evolution (Kino *et al.,* 2004, Towers *et al.*, 2003, Sheehy *et al.*, 2003, Kobayashi *et al.,* 2004).

Numerous approaches have been used for the functional genomics analysis of the effects of HIV on host cells, using microarrays. These include infection with HIV produced from infectious molecular clones or primary isolates (van 't Wout *et al.,* 2003, Vahey *et al.,* 2002, Geiss *et al.,* 2000) as well as the study of HIV-1 accessory genes/proteins in isolation (Izmailova *et al.,* 2003). In the latter example it was found that HIV-1 infection or Tat expression induces interferon-responsive gene expression in immature human dendritic cells without inducing maturation. Among the induced gene products are chemokines that recruit activated T cells and macrophages, the ultimate target cells for the virus, suggesting that HIV-1 Tat reprograms host dendritic cell gene expression to facilitate expansion of HIV-1 infection. To address the function of the HIV-1 accessory protein, Vpr, cDNA microarray techniques were used to analyse the regulation of a panel of host cellular genes during infection with isogenic HIV-1; either with or without Vpr expression. Results indicate that Vpr can down regulate the expression of genes involved in cell cycle/proliferation regulation, DNA repair, tumour antigens and immune activation factors, and up regulate many ribosomal and structural proteins. Such information regarding the involvement of several cellular genes in HIV infection is thus very important, both for understanding Vpr functions and for the

development of therapeutics targeting the Vpr molecule (Janket *et al.*, 2004). Cell lines engineered to constitutively express HIV accessory proteins including Tat and Nef have also revealed much about their functional interactions with the intracellular milieu (Shaheduzzaman *et al.*, 2002, de la Fuente *et al.*, 2002). In Nef expressing cell lines genes expressed at higher levels include proteases, transcription factors, protein kinases, nuclear import/export proteins, adaptor molecules and cyclins, some of which have previously been implicated as being important for HIV replication and pathogenesis. These data indicate that Nef expression can alter the expression of cellular genes and suggest that this alteration in cellular gene expression may serve to optimise the cell to support subsequent stages of viral replication (Shaheduzzaman *et al.*, 2002). Conversely most of the cellular genes in Tat expressing cells were down regulated. This was most apparent for cellular receptors that have intrinsic receptor tyrosine kinase (RTK) activity and signal transduction factors that mediate RTK function, including the Ras-Raf-MEK pathway (de la Fuente *et al.*, 2002). The authors postulate that this down regulation of receptors may allow latent HIV-1 infected cells to either hide from the immune system or avoid extracellular differentiation signals. A small number of genes were found to be up regulated, however, including co-receptors for HIV-1 entry, translation machinery, and cell cycle regulatory proteins (de la Fuente *et al.*, 2002).

By comparing gene expression in cells both permissive and resistant to HIV infection (Kartvelishvili *et al.*, 2004), patterns have also been characterised that correlate with resistance. Specifically, analysis of a cell clone resistant to HIV-1 infection revealed that it expressed a soluble resistance factor and displayed changes in several cellular genes affecting HIV-1 susceptibility and expression. Among up regulated genes, three were linked directly to the cellular resistance to HIV-1, one being the transcription repressor CTCF. The remaining two are yet to be characterised.

Moving closer to the *in vivo* situation, a unique study published in 2004 used functional genomics to evaluate the effects of antiretroviral therapy on HIV-1 infection of lymphatic tissue (Li *et al.*, 2004a). By studying patients both before and after treatment in cross section, approximately 200 treatment-responsive genes - many of them known mediators and moderators of immune activation and defences, particularly innate defence genes - were found to be expressed at all stages of HIV-1 infection. Most of the treatment-responsive genes were categorised as mediators of trafficking, reformation of

active follicles and tissue repair. The authors propose a model in which nearly counterbalanced functions of mediators and moderators of immune activation and defences account for the slow dynamics of HIV-1 infection before treatment, suggesting that there could be a role for anti-inflammatory agents, alone or in combination with HAART, in treating HIV-1 infection by altering this balance to mitigate pathology. A similar, but much smaller study, looked at the restoration of gut associated lymphoid tissue (GALT) after onset of HAART (Guadalupe *et al.,* 2003). An increase in expression of genes associated with cell trafficking was observed (supported by other experimental data) demonstrating that lymphocyte homing plays an important role in the restoration of intestinal CD4+ T cells in GALT following HAART. An additional *in vivo* study looked at latently infected, resting CD4+ T-cells of HIV positive individuals (Chun *et al.,* 2003) during phases of viraemia and aviraemia. It was found that these cells, from viraemic patients, are capable of producing cell-free virus spontaneously *ex vivo,* whereas those from aviraemic patients (despite evidence for transcription of RNA) did not. Expression profiling of these resting T-cells from viraemic patients revealed specific and significant up regulation of genes involved in transcription regulation, RNA processing and modification, protein trafficking and vesicle transport, suggesting that active viral replication has a profound effect on the physiologic state of resting CD4+ T-cells, ultimately resulting in the release of HIV without exogenous stimuli. Finally, animal models have also been used to further understand the pathogenesis of neuro-AIDS (Roberts *et al.,* 2003) and thymic disruption (Miller *et al.,* 2003).

To date, however, no microarray studies have addressed the effects of different genetic variants of HIV on host cell gene expression. As described in earlier chapters, the full range of HIV genetic variants are of clinical significance in that they play a role in the HIV-1 epidemic. Furthermore, HIV isolates representing the considerable genetic variation within HIV-1 have quite different growth characteristics on the same T-cell substrate. The hypothesis may therefore be put that these differences might be further explained by identifying differences in the host cell transcriptome during infection with different strains of HIV-1. Differential modulation of host gene expression may provide evidence in support of the published differences in pathogenicity, disease progression or transmissibility of different viral subtypes, groups or at the furthest ends of the spectrum, between HIV-1 and HIV-2. Microarray profiling of cells infected with different HIV-1 subtypes and HIV-2 may also identify a 'core' HIV response program.

molecular mechanism of pathogenesis of the related yet quite distinct hantaviruses Sin Nombre and Prospect Hill (Khaiboullina *et al.*, 2004). Finally, array based comparisons of hepatitis B and C viruses have revealed the mechanisms by which these very different viruses cause the same clinical endpoint, hepatocellular carcinoma (HCC). Characterisation of these distinct mechanisms potentially provides a novel tool for diagnosis and treatment of HBV- and HCV-associated HCCs (Okabe *et al.*, 2001, Iizuka *et al.*, 2002).

In the case of HIV, however, such array-based studies comparing genetic variants of the same viral species in an attempt to explain their phenotypic differences have not been undertaken. This is perhaps not surprising, as it is widely believed that the heterogenous distribution of different HIV-1 subtypes may not be due to any inherent properties of the viral strains themselves but rather the chance results of founder effect where one or more subtypes are introduced and consequently established in a population before other subtypes. These subtypes have subsequently been classified by modern phylogenetic techniques (Robertson *et al.*, 2000). No large scale, accurate subtyping, however, has been carried out on infected populations in order to facilitate ongoing epidemiological studies of the relationship between HIV subtype and disease (Chapter 2). Because the subtype classification of HIV isolates has not come about as a result of detectable, specific biological or phenotypic characteristics, it follows that functional genomics has not been immediately exploited to further understand the importance of HIV subtypes. Epidemiological evidence, *in vitro* studies and the observations described in this thesis, however, have suggested that there may be differences between biological properties of different HIV genotypes. Microarray studies might therefore be of use in trying to understand the impact of HIV sequence variation on viral and therefore perhaps disease phenotype.

4.1.5 Using functional genomics in the study of HIV subtypes

The following experiments describe the effect of HIV-1 subtype B, HIV-1 group O and HIV-2 on the transcriptional profiles of the CD4+ T-cell line ST1-R5. In chapter 2 the growth characteristics of these viruses was established on different T-cell lines and these isolates provide what are perhaps the ends of the spectrum of terms of growth phenotype. The inclusion of HIV-2 is also pertinent, as clear differences between HIV-

2 and HIV-1 have been shown, both in terms of disease characteristics and their direct effect on cells *in vitro*.

4.1.5.1 HIV infection of T-cells

The first high-density microarray studies of HIV infection of T-cells described the transcriptional effects of low multiplicity HIV-1 LAI infection on the CD4+ T-cell line CEM-CCRF (Geiss *et al.*, 2000) and human PBMCs (Vahey *et al.*, 2002), in time course assays. These studies revealed that multiple genes could be identified that were specifically and consistently up and down regulated during HIV infection. The results of both studies, however, are difficult to interpret. Both infection at low multiplicity (leading to an asynchronous infection) (Geiss *et al.*, 2000) and studying a heterogenous PBMC target cell population (Vahey *et al.*, 200) whilst examining cellular gene expression 2-3 days post-infection can lead to confounding changes in gene expression over time.

More recently these concerns were addressed by the use of high titre virus to infect various CD4+ T-cell lines, monitoring the level of >4600 cellular RNA transcripts within 24 hours of infection (van 't Wout *et al.*, 2003). A total of 409 of 4608 genes were significantly regulated at one or more time points post-infection: 165 had increased and 244 had decreased levels of stable RNA expression. Overall levels of signalling and activation were found to be increased, including several genes involved in cell activation, signalling and communication that were up regulated within 1-8 hours post infection, prior to detection of significant virus replication. This observation was also made in cells infected at a lower MOI (Geiss *et al.*, 2000, Vahey *et al.*, 2002) or when exposed to HIV-1 Nef (Simmons *et al.*, 2001). In addition infected cells had decreased levels of proliferation, reflected in a decrease in the expression of cell cycle inducers and an increase in expression of repressors. This was proposed to reflect the cell-cycle G2 arrest generated by HIV-1 in response to viral Vpr (Planelles *et al.*, 1996). Genes involved in cell division and transcription, a family of DEAD box protein RNA helicases, and genes involved with translation and RNA splicing were mainly down regulated. A small group of transcription factors including EGR-1 and JUN, however, were up regulated, alluding to a specific role of these in increasing HIV-1 production. Finally, a series of enzymes involved in cholesterol biosynthesis, regulated by the sterol

responsive element binding protein 2 (SREPB2), were up regulated in all cell lines studied (van 't Wout *et al.*, 2003).

In this seminal study, it was found that transcriptional changes induced by HIV-1 were most pronounced at 24 hours post-infection. At this stage regulation differences were also evident between different cell types studied, prompting the suggestion that studying host-cell responses in different target cells (i.e. both permissive and non-permissive) might provide an opportunity to identify host proteins essential for HIV-1 replication (van 't Wout *et al.*, 2003).

Given the findings of previous microarray experiments looking at HIV infection of T-cells, it was decided that a similar approach should be taken in that infection of T-cells with HIV-1 subtype B, group O and HIV-2 should be carried out at a high multiplicity and gene expression should be analysed within 24 hours of infection. As described in Chapter 2, the infectious titre of the primary isolate stocks (as determined by NP2 cell titration) is low compared to that which can be achieved with cell line adapted viruses or molecular clones. To achieve the desired high multiplicity of infection a small number of cells must therefore be infected. When RNA is limiting in microarray experiments, an amplification step to generate sufficient RNA/cDNA for use with a standard labelling method may be used. Furthermore, with very small cell numbers, extracting mRNA prior to amplification also becomes problematic and requires careful optimisation.

4.1.5.2 RNA amplification

Amplification of RNA was developed and used in gene expression analysis long before the advent of microarray technology (Van Gelder *et al.*, 1990). This involved reverse transcription of mRNA with an oligo dT/T7 promoter primer, followed by synthesis of double stranded cDNA, antisense RNA then transcribed *in vitro* using T7 polymerase generally resulting in a 1000-fold amplification of the original amount. Unlike PCR amplification this method was found to be linear and not biased by the size of the template (under optimal conditions). It is the case that the majority of amplification methods used today in microarray experiments are based on this methodology, and many publications have addressed the question of reproducibility of different amplification methods and attempted to quantify the bias introduced to the data by the

RNA amplification process. The main conclusions from such work are that, although amplification methods introduce slight changes in the transcript ratios compared to standard labelling, they are highly reproducible and, for small sample size, *in vitro* transcription is the preferred method. Different labelling strategies, however, should not be used within a single study, and all samples should be equally treated (Gomes *et al.*, 2003, Jenson *et al.*, 2003, Schneider *et al.*, 2004, Wilson *et al.*, 2004). This includes the length of time the RNA is amplified for, as time-dependent degradation of aRNA has been observed with amplifications exceeding 4 hours (Spiess *et al.*, 2003). Thus, there are certain "quality vs. yield" issues inherent in T7-based RNA amplification. It has also been proposed that there should be a general standardisation of the RT reaction for small-sample profiling with regard to the Minimal Information about a Microarray Experiment (MIAME) requirements (Kenzelmann *et al.*, 2004). A recent report that total RNA, mRNA and aRNA can all be used in expression profiling analysis as long as the test and reference samples are generated by identical method within single study is also important to note. That is, if using amplified RNA the reference must also be amplified (Li *et al.*, 2004b).

4.1.6 Aims

The aim of this chapter is to use cDNA microarrays to determine whether genetically distinct strains of HIV result in differing modulation of host transcription, during the first 24 hours of infection. The consequences of this in terms of viral phenotype, disease phenotype, biomarkers of disease and identification of potential novel therapeutic targets will be discussed.

4.2 Methods

4.2.1 Cell and viruses

The viruses used were those previously characterised (see Chapter 3): HIV-1 subtype B (SF162), HIV-1 group O (BCF06) and HIV-2 (ETP). The cell line used for these experiments was ST1-R5 (mycoplasma free), which had been previously used to study these isolates. The ST1-R5 cells were maintained in RPMI 1640-2 mM L-glutamine medium, supplemented with 10% FBS, 100 U/ml penicillin, 100 μg/ml streptomycin and 1 μg of puromycin/ml.

4.2.2 Reference RNA

For optimising and assessing RNA amplification methods, RNA was extracted from large numbers of ST1-R5 cells.

4.2.2.1 RNA extraction

10^6 ST1-R5 cells were centrifuged in individual 15 ml centrifuge tubes, resuspended in 1.0 ml TRIZOL (Gibco) and stored at -80°C. For extraction, frozen lysates were thawed at 37°C and centrifuged at 12,000 xg for 10 minutes at 4°C. Following incubation at room temperature (RT) for 5 minutes, 250 μl chloroform was added and the tube shaken for 15 seconds, incubated for a further 3 minutes at RT then centrifuged at 12,000 xg for 15 minutes (4°C). The aqueous phase was transferred to a new tube containing 625 μl chloroform and shaken for 15 seconds, incubated at RT for 3 minutes, then centrifuged at 12,000 xg for 15 minutes (4°C) and the aqueous phase transferred to a new tube. 625 μl isopropanol was added, the tube vortexed and incubated at RT for 10 minutes, followed by centrifugation at 12,000 xg for 15 minutes (4°C). The pellet was washed in 1ml 75% ethanol and air-dried at RT for 5 minutes before re-suspending in 50μl RNase-free water.

4.2.2.2 DNaseI treatment and purification of RNA

After RNA extraction, the samples were treated with DNaseI (Promega) in a final volume of 200 μl, at 37°C for 1 hour (h), after which terminator solution was added ($1/10^{th}$ total volume) (0.1 M EDTA, 1 μg/μl glycogen) to stop the reaction. An equal volume of phenol:chloroform:isoamyl alcohol (Sigma) was added to the DNase treated RNA, vortexed for 10 seconds then centrifuged at 13,000 rpm for 10 minutes at RT.

The top aqueous layer was transferred to a new tube and phenol:chloroform extraction was repeated. The RNA was precipitated using 1/5th volume 8 M ammonium acetate and 2.5 volumes 95% ethanol at -20°C for 2 h, then centrifuged at 14,000 rpm for 30 minutes (4°C) to pellet. The RNA was washed with 200 µl of ice-cold 80% ethanol and air dried for 5 minutes, prior to re-suspension in 30 µl DEPC-water and immediate storage at -80°C.

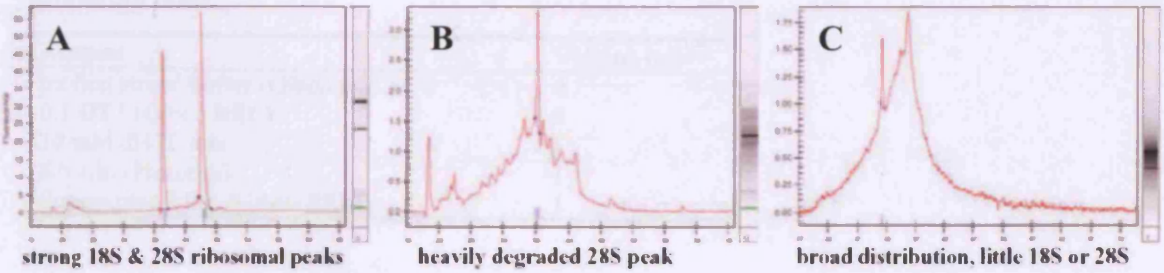
4.2.2.3 mRNA extraction

Two methods of mRNA extraction were used from total RNA extracts or cell lysates: Oligotex mRNA mini kit (Qiagen) and Dynal Dynabeads (Dynal Biotech, Sweden). The reaction mixtures were adjusted according to the total amount of RNA that had been extracted and manufacturer's protocol followed. In short the oligotex protocol relies upon the selective binding of poly-A mRNA to beads coated with oligo-dT oligomers. The method involves several wash steps using a column and filter system, the mRNA eventually eluted from beads requiring subsequent concentration, achieved using a Microcon-30 filter in a Microcon 1.5ml tube (Millipore). Similarly, the Dynal beads system relies upon the oligo-dT binding capacity of mRNA, in this case however the coated beads are magnetic and can be drawn out of solution using a magnet. After washing steps, the mRNA may be eluted from the beads into as small a volume as desired. Consequently, mRNA concentration is not necessary. For both methods RNA was eluted into RNase free DEPC-water and immediately stored at -80°C.

4.2.2.4 RNA quantitation

Both total RNA (TRNA) and mRNA were quantified using an Agilent Bioanalyser, which detects RNA by laser-induced fluorescence. Specifically, each Nano-LabChip® contains an interconnected set of microchannels that separates nucleic acid fragments based on their size as they are driven through it electrophoretically. Each chip is prepared by addition of a gel-dye mix followed by individual RNA samples (up to 12 per chip) to each well. RNA ladder and marker (RNA 6000 ladder, Ambion, USA) are added to separate wells. During the chip run, the dye intercalates directly with the RNA and all bands pass the detector at different speeds. The software automatically compares the unknown samples to the ladder fragments to determine the concentration of the unknown samples

Figure 4.1 Representative bioanalyser traces.
(A) good quality TRNA; **(B)** degraded TRNA and **(C)** good quality mRNA (source: http://www.agilent.com).



and to identify the ribosomal RNA peaks. The ladder also serves as a built-in quality control measurement of system performance under standard conditions. Resulting, therefore, is an assessment of both quality and quantity of RNA (see Figure 4.1 for example bioanalyser output) (Naderi *et al.*, 2004). The RNA 6000 Nano LabChip Kit was used, as samples analysed were all expected to be in the range of this assay (5-500 ng/µl for TRNA and 25-250 ng/µl for mRNA/aRNA). All reagents were allowed to warm to RT for 30 minutes and RNA ladder was denatured at 65°C for 4 minutes then held at 4°C, prior to use.

4.2.3 RNA amplification

Ways to both extract and amplify RNA were investigated, in order to determine the optimal conditions for preparing mRNA for microarray analysis.

4.2.3.1 In vitro amplification of RNA - In house

A T7-based method of RNA amplification was previously developed in our laboratory. It is a modified Eberwine method (Van Gelder *et al.*, 1990) and was used to amplify from ST1-R5 TRNA (extracted using Trizol). The following were mixed in a 0.5 ml PCR reaction tube:

Reagent	Volume (µl)
Total RNA (1µg)	10
OligodT-T7 primer (0.5 µg/µl)	1

OligodT(21)-T7 primer: 5' TCT AGT CGA CGG CCA GTG AAT TGT AAT ACG ACT CAC TAT AGG GCG T(21) 3' [56-mer]

These were placed at 70°C for 10 minutes, chilled on ice followed by incubation at 42°C for 5 minutes. The first strand master mixed was prepared according to the following table:

Reagent	Volume (µl)
5x first strand buffer (Gibco BRL)	4
0.1 DTT (Gibco BRL)	2
10 mM dNTP mix	1
RNasin (Promega)	1
Superscript RT II (Gibco BRL)	1

The mastermix was added to the samples and incubated at 42°C for 1 h. 1µl was removed for PCR (see table below) to check whether intact RNA was successfully obtained.

Reagent	Volume (µl)
ss cDNA	1
10x PCR buffer (Qiagen)	5
10 mM dNTPs (Promega)	1
Forward primer (100 ng/µl)	2
Reverse primer (100 ng/µl)	2
Taq DNA polymerase (5 U) (Qiagen)	0.5
Sterile water	38.5

The second strand master mix was subsequently prepared:

Reagent	Volume (µl)
Second strand synthesis buffer (Gibco BRL)	30
10 mM dNTPs	3
DNA poll (Gibco BRL)	4
E.Coli RnaseH (Gibco BRL)	1
E.Coli DNA ligase (Gibco BRL)	1
DEPC water	92

The master mix was added to the remainder (19 µl) of the first strand reaction and incubated at 16°C for 2 h. 2 µl of T4 DNA polymerase (Gibco) were added and samples incubated for a further 10 minutes at 16°C. The samples were extracted once with phenol/chloroform/isoamyl alcohol (25:24:1) (Sigma) and the aqueous layer was collected and purified using a pre-rinsed Microcon-100 column (Millipore) according to manufacturer's instructions. If necessary, the sample volume was reduced to 8 µl by vacuum centrifugation.

The Ampliscribe T7 transcription kit (Cambio) was used for *in vitro* transcription. Specifically, the transcription mix was prepared adding the reagents in the following order:

Reagent	Volume (µl)
10xAmp T7 buffer	2
ATP	1.5
CTP	1.5
GTP	1.5
UTP	1.5
0.1 M DTT	2

These were added to the 8 µl of ds cDNA along with 2 µl of T7 RNA polymerase and incubated at 42°C for 3 h. 1 µl of RNase-free DNase was added and reaction placed at 37°C for 15 minutes. The aRNA was purified as before and stored at -80°C. Quality was assessed using an RNA 6000 Nano LabChip Kit.

4.2.3.2 In vitro amplification of RNA – RiboAmp RNA Amplication kit

The second RNA amplification protocol used was also a T7-based method, but with inclusion of an internal primer for amplification from both ends of synthesised cDNAs (RiboAmp RNA amplification kit, Arcturus, UK). Unlike the in-house method for amplification, this protocol is optimised for amplification of very small amounts of RNA and is designed to be used with the Picopure RNA Isolation Kit (Arcturus, UK), which isolates TRNA, prior to amplification. Both methods were followed as per manufacturer's instructions. The effect of using Dynal Dynabeads to extract mRNA from cells prior to RiboAmp amplification was also assessed, as was using different starting cell numbers (from which RNA was extracted). RNA quality was assessed using an RNA 6000 Nano LabChip Kit.

4.2.3.3 PCR screening of cDNA for GapDH

To assess the efficiency of cDNA synthesis following RNA extraction by different methods (prior to amplification), cDNA was screened for presence of glyceraldehyde-3-phosphate dehydrogenase (GapDH). This is a cellular gene whose expression is generally maintained at a detectable level throughout the life of a cell, therefore providing a good positive control. The constituents of each reaction were:

Reagent	Volume (µl)
DNA (approx. 50 ng)	2
Primers (10 pMol/µl)	1 each
PCR buffer (10x)	5
$MgCl_2$ (25 mM)	4
dNTPs (10 mM)	1
Taq DNA polymerase (5 U/µl)	0.25
Distilled water	35.75

Cycling conditions were as in section 3.2.7.1. PCR products were separated on a 1% w/v agarose gel in 1x TBE containing 0.2 μg/ml ethidium bromide (Sigma, UK) at 80V.

4.2.4 Time courses

Cells were infected at an MOI of 1.0, the multiplicity calculated as the ratio of infectious units (as defined using NP2 cell titration) to cell number. Specifically, for each infection 10^4 cells per time point were incubated with the appropriate volume of virus at 37°C for 3 h. The volume was equalised for each infection. Cells were then centrifuged to pellet and washed once in medium prior to addition to a 96 well plate (10^4 cells per well), and incubation at 37°C. Two time courses were carried out, one during which samples were harvested at 0 and 24 hours post-infection (HPI) and a second where 0, 2, 6, 12 and 24 HPI were sampled. Each time point was performed in triplicate. Given prior investigation, the Picopure total RNA isolation kit was used to extract RNA from samples and mRNA was subsequently amplified using the RiboAmp RNA amplification kit. Success of amplification was determined using an RNA 6000 Nano LabChip Kit.

4.2.5 Labelling

A direct incorporation method (CyScribe First-Strand cDNA Labelling Kit) was used (Amersham Biosciences, GE HealthCare). Specifically, the following reagents were added to 1 μg of amplified RNA, 500 ng of mRNA or 10 μg TRNA in a 0.5 ml PCR tube on ice: 1 μl random nonamers (for aRNA), 1 μl anchored oligo(dT) (for mRNA or TRNA) and water to a final volume of 11 μl. The reaction mixture was incubated at 70°C for 5 minutes followed by incubation at RT for 10 minutes. Subsequently, 4 μl 5x CyScript buffer, 2 μl 0.1M DTT, 1 μl dCTP nucleotide mix, 1 μl dCTP CyDye-labelled nucleotide and 1 μl CyScript reverse transcriptase were added to the reaction mixtures. These were centrifuged at 13,000 rpm at RT for 30 seconds and then incubated at 42°C for 1.5 h. On all occasions reference RNA was labelled with Cy3 labelled nucleotide and sample with Cy5. The RNA was then denatured by addition of 2.5 μl 0.5 M EDTA (pH 8.0) and 10 μl 0.1 M NaOH and incubated at 70°C for 10 minutes, the reaction neutralised by addition of 10 μl 0.1 M HCl, 3 μl CoT-1 DNA (Gibco), and 450 μl TE

(pH 8.0). Labelled RNA was purified using a Microcon-100 filter and reduced to 12 µl. The labelled RNA was immediately prepared for hybridisation.

4.2.6 Hybridisation

The Cy3 and Cy5 labelled probes were made up to 14 µl with TE (pH 8.0) and the hybridisation mixtures prepared follows: 12 µl 20x SSPE (Sigma), 1.1 µl 0.5 M EDTA, 2 µl poly $dA_{40\text{-}60}$ (Pharmacia), 2 µl yeast tRNA (Sigma), 14 µl Cy3 probe, 14 µl Cy5 probe, and finally 1 µl 10% SDS. The mixture was vortexed and incubated at 98°C for 2 minutes followed by 37°C for 20 minutes. Subsequently 1µl 100x Denhardt's solution (Sigma) was added to the probe mixture and the mix centrifuged at 13,000 rpm for 15 minutes at RT. The human cDNA Gen2 arrays (Hs_clone_Av2, MRC-HGMP, UK) were transferred to hybridisation chambers, pre-heated to 65 °C, and allowed to warm for 30 minutes. Glass cover slips were cleaned by immersion in 95% ethanol for 2 minutes and centrifugation at 1000 rpm for 2 minutes to dry. The probe (46 µl) was applied to the array by adding it to the glass coverslip, this was then applied to the pre-warmed array by gently lowering the array over the coverslip until capillary action adhered the two together and distributed the probe, with no need for application of pressure. The array was placed back into the hybridisation chamber and 150 µl 4x SSPE (65°C) added, the chamber lid was firmly secured and the assembly transferred to a 65°C water bath and incubated for 16 h.

After incubation the hybridisation chamber was dismantled at 65°C and the array carefully removed and immersed in 2x SSPE heated to 50°C. Once the coverslip had dissociated the array was immersed in fresh 2x SSPE for 2 minutes at RT (rolling), followed by 1x SSPE for 2 minutes and 0.1x SSPE for 3 minutes. Subsequently, the array was transferring it to a clean 50 ml Falcon tube and centrifuged at 1,000 rpm for 2 minutes to dry.

4.2.7 Array scanning

Arrays were scanned at 10 µm resolution using the GenePix 4000B array scanner (Axon instruments, Molecular Dynamics) and the images analysed using GenePix Pro 3.0 software. Cy3 and Cy5 fluorophores were simultaneously excited at 532 nm and 635 nm respectively and the resultant light detected with two PMTs. The voltages across the

PMTs were adjusted so that the signals from the two array channels were balanced. The GenePix software combines the data from the two channels to create a single composite image. A template was fitted over the array image using a spot-finding software algorithm. All the elements on each array were checked by eye and manually corrected where necessary. Data were extracted from the image by the software using the adjusted template and the normalisation factor (average ratio between signals in the Cy3 and Cy5 channels) calculated automatically by the GenePix software. Expression ratios were calculated as the median of the ratios between the local background-subtracted Cy3 and Cy5 signals on a pixel-by-pixel basis by the software. The data were exported to a spreadsheet created in Excel, ArrayAnalyser. The median of ratios were filtered to remove flagged array elements and elements for which the signal to background ratio was below that calculated as the average for negative control elements included on the array, in both Cy5 and Cy3 channels.

4.2.8 Array analysis

4.2.8.1 Cluster and treeview

The final array data were analysed using Cluster software (Eisen *et al.*, 1998). The program Cluster assembles a set of elements (genes or arrays) into a tree, where elements are joined by tree branches whose length is proportional to the distance (correlation coefficient) between the elements. Array elements for which expression measurements failed the signal to noise ratio filter were removed and the data converted to log(2). The arrays and genes were then median centred (the median expression ratio within each array or of each array element across all arrays was set to 0). Genes and/or arrays were clustered by average-linkage hierarchical clustering and a self-organising map algorithm applied, in order to direct the orientation of nodes generated by hierarchical clustering. The results were visualised using the software Treeview.

4.2.8.2 Significance analysis of microarrays

SAM (Significance Analysis of Microarrays) is a statistical technique for finding significantly differently expressed genes in a set of microarray experiments (Tusher *et al.*, 2001) and functions as a software package (written by B. Narasimhan) for Excel. The input to SAM is a gene expression matrix from a set of microarray experiments, as well as a response variable from each experiment. In the case of these experiments, the

response variable is infected at time X vs. infected at time Y, or infected with virus X vs. infected with virus Y. By inputting normalised log2 ratios (>4 replicates for each array/response variable) SAM identifies genes with statistically significant changes in expression by assimilating a set of gene-specific T-tests. Each gene is assigned a score on the basis of its change in gene expression relative to the standard deviation of repeated measurements for that gene. Genes with scores greater than a threshold are deemed differentially expressed. The percentage of such genes identified by chance is calculated by a false discovery rate (FDR). To estimate the FDR, genes are randomly permuted in a form of bootstrapping. The FDR threshold can be adjusted to identify smaller or larger sets of genes. The FDR for these experiments was selected as 5%.

4.2.8.3 Gene Annotation

Genes were annotated using DAVID (Database for Annotation, Visualization and Integrated Discovery, http://apps1.niaid.nih.gov/david/). By inputting a list of accession numbers Gene Ontology (GO) pathways were produced, which provide a unique identifier for each gene annotated for a cellular process or function within GO. Genes were then matched to the KEGG (Kyoto Encyclopedia of Genes and Genomes) description of known cellular pathways.

4.3 Results

In order to perform microarray studies on HIV infection of T-cells, a small number of cells were infected at an MOI of 1.0, with HIV-1 subtype B, group O and HIV-2 ETP. Accordingly a number of protocols had to be performed and optimised, including extraction of RNA, synthesis of reference RNA, analysis of amplification methodologies and optimisation of time course strategies, prior to the microarray experiments themselves.

4.3.1 RNA extraction and quality

4.3.1.1 RNA extraction and amplification

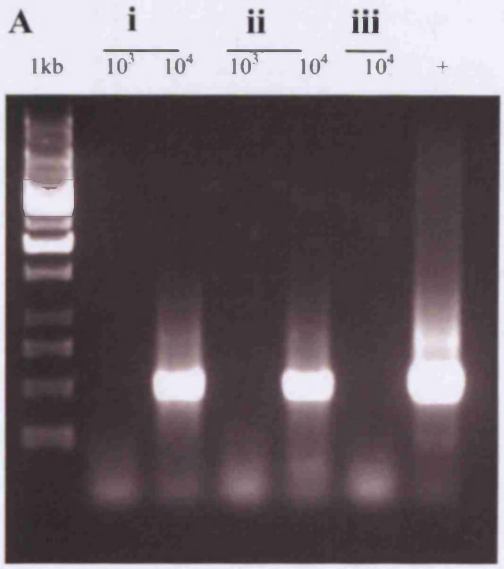
Although a variety of methods have been developed for the extraction and purification of RNA, prior to amplification or labeling for microarray analysis, there have been few systematic evaluations to optimise these methods. A recent publication went some way to address this problem by experimentally determining the best method for each step of RNA extraction with respect to yield, purity and size distribution of the transcripts. In short, DNase treatment of diluted total RNA samples followed by phenol extraction is the optimal way to remove genomic DNA contamination. Purification of double stranded cDNA is best achieved by phenol extraction followed by isopropanol precipitation at room temperature. Extraction with guanidinium-phenol and lithium chloride precipitation are the optimal methods for purification of amplified RNA (aRNA) and labelled aRNA respectively (Naderi *et al.*, 2004). Whilst demonstrating the importance of an optimised RNA extraction protocol, this study did not address how best to extract RNA from small sample sizes. This problem has to a large extent been overcome by the commercial development of several RNA extraction methods designed for samples as small as a single cell. These methods usually take two forms in that they either enable extraction of total RNA (TRNA) (usually using membrane/filter technologies) or mRNA (relying on poly-dT coated resin and filtration, or magnetic beads). How three of these different methods affect RNA quality, the most critical factor for microarray success, was assessed here. The methods are Trizol extraction, Dynal Dynabeads (mRNA, Dynal) and Oligotex (mRNA, Qiagen).

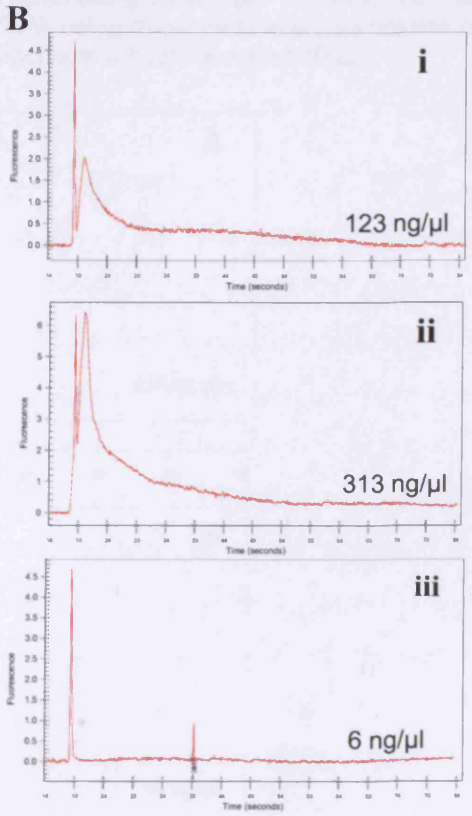
Specifically, the three RNA extraction methodologies were tested on 10^3 and 10^4 SupT1-R5 (ST1-R5) cells in order to assess their efficiency for processing small sample sizes. Extractions were performed as described and the efficiency of extraction assessed by performing cDNA synthesis on extracted RNA followed by PCR screening for glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Results showed that RNA could not be effectively isolated from 10^3 cells by any method, as GAPDH could not be detected in cDNA synthesised from extracted TRNA (using Trizol) or mRNA (Dynabeads and Oligotex). Using 10^4 cells, however, GAPDH was clearly detectable in cDNA synthesised from Oligotex- and Dynabeads-extracted mRNA, but not from Trizol extracted TRNA, implying 10^4 cells is a reasonable number to use if the former two methods are to be employed (Figure 4.2-A). To extend this analysis, 2 rounds of amplification were performed from the extracted mRNA using an in-house modified Eberwine protocol previously validated. These data indicate that, after extracting RNA from 10^4 cells using both Oligotex and Dynabeads, efficient amplification can be achieved (Figure 4.2-B), and that the aRNA yield is higher when amplifying from mRNA extracted with Dynabeads. After extraction of TRNA from 10^4 cells using Trizol, however, no amplification was possible. Taken together, these data indicate that of the three methods, Trizol should not be used for extracting RNA from small samples prior to amplification and, whilst Dynabeads and Oligotex give reasonable results from 10^4 cells, they cannot be used to successfully extract mRNA (in order to synthesise cDNA) from 10^3 ST1-R5 cells.

An additional observation, with respect to the quality of amplified RNA (Figure 4.2-B (i) and (ii)) is that, irrespective of the RNA extraction method, the majority of aRNA produced is very short in length; ranging from 80 – 90 nucleotides. It has previously been reported that the size distribution of aRNA is strongly dependent on quantity of input RNA. For example, when the RNA input is below 50 ng of mRNA or 100 ng of TRNA, the average size of the aRNA after one round of Eberwine amplification is less than 200 nucleotides (Li *et al.,* 2003), in accordance with the results produced here. Current advances in amplification technology, however, mean that several techniques now exist that do not reduce aRNA length to this extent.

Figure 4.2 **Evaluation of extraction and amplification methods.**

Analysis of the efficiency of:
(**A**) reverse transcription and GapDH amplification from 10^3 or 10^4 ST1-R5 cells, and (**B**) result of two rounds of amplification of RNA, extracted by three different methods (i) Oligotex; (ii) Dynabeads; (iii) Trizol, from 10^4 ST1-R5 cells.



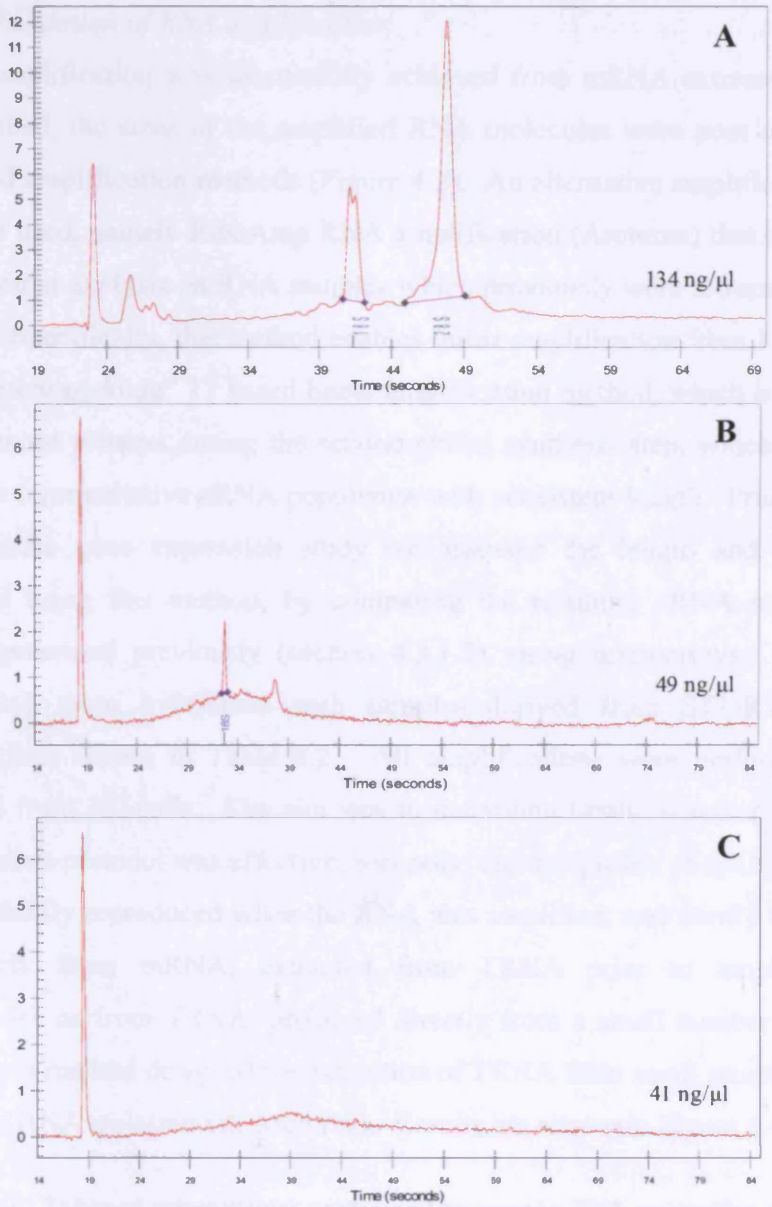


It was therefore decided that an alternative amplification method should be optimised, working from 10^4 cells with the aim of improving the quality of aRNA produced. As both Oligotex and Dynabeads gave similar results in terms of efficiency of extraction and subsequent amplification, it was decided that both methods would be used to produce mRNA reference, and a substrate on which amplification methods could be tested, the aim being to select the best combination of RNA extraction method and amplification procedure.

4.3.1.2 Synthesis of reference RNA

TRNA was extracted from 10^6 ST1-R5 T-cells to both provide a control for validation of amplification techniques using microarrays, and from which to extract mRNA (for amplification). This was achieved using a Trizol extraction method followed by mRNA extraction using either Oligotex or Dynabeads, in the latter two cases 90 µg of TRNA used for extraction of mRNA. The quality and quantity of isolated RNA was assessed using an Agilent bioanalyser (Figure 4.3).

Figure 4.3 Bioanalyser traces of ST1-R5 TRNA and mRNA.
Representative bioanalyser traces for: (A) TRNA, extracted using Trizol from 10^6 cells, in a final volume of 100 µl; (B) mRNA, extracted from 90 µg TRNA using dynabeads in a final volume of 40 µl; (C) mRNA, extracted from 90 µg TRNA using oligotex in a final volume of 40 µl.



Extraction of TRNA using the Trizol method was effective and yielded RNA, with no observable DNA contamination and no degradation (Figure 4.3-A). mRNA extraction was achieved using two separate methods, both of which yielded mRNA of approximately the same quality and concentration (Figure 4.3-B and C). Assuming that approximately 2% of the TRNA used for these extractions should be mRNA (http://www.ambion.com/techlib/tn/93/9313.html) the yield achieved was acceptable.

That is, in total 1640 to 1960 ng of mRNA was extracted from 90 μg TRNA, approximately 1.8 – 2.2 %.

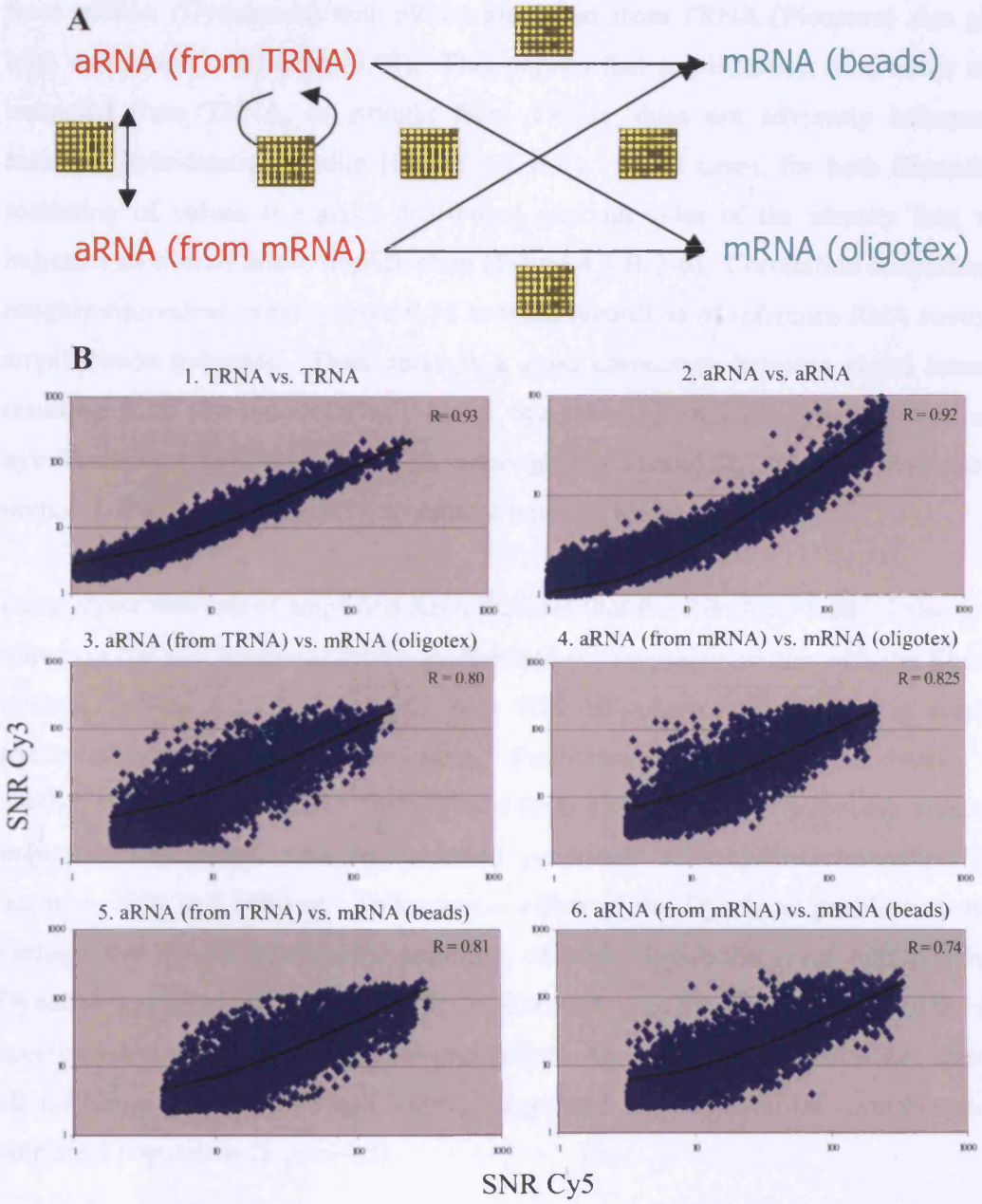
4.3.1.3 Validation of RNA amplification

Whilst amplification was successfully achieved from mRNA extracted from 10^4 cells, as described, the sizes of the amplified RNA molecules were poor compared to other published amplification methods (Figure 4.2). An alternative amplification method was therefore used, namely RiboAmp RNA amplification (Arcturus) that has been designed for molecular analysis on RNA samples which previously were too small for microarray analysis. Specifically, this method enables linear amplification from 1 ng of TRNA and has a "patent-pending" T7 based linear amplification method, which is distinct in its use of exogenous primers during the second strand synthesis step, which are optimised to provide a representative aRNA population with consistent length. Prior to commencing a large-scale gene expression study we assessed the length and quality of RNA amplified using this method, by comparing the resulting aRNA to the TRNA and mRNA generated previously (section 4.3.1.2), using microarrays. Specifically, six microarrays were hybridised with samples derived from ST1-R5 TRNA in the combinations shown in Table 4.2. All amplifications were performed using RNA extracted from 10^4 cells. The aim was to determine firstly, whether the labelling and hybridisation protocol was effective; secondly, that the profile of mRNAs within TRNA were faithfully reproduced when the RNA was amplified; and thirdly whether it is best to amplify from mRNA, extracted from TRNA prior to amplification (using Dynabeads), or from TRNA, produced directly from a small number of cells. In the latter case a method designed for extraction of TRNA from small sample sizes was used (Picopure RNA isolation kit, Arcturus). Results are shown in Figure 4.4.

Table 4.2 Table of microarrays performed to assess RNA extraction and amplification methodologies.

Cy3		Cy5
ST1-R5 TRNA		ST1-R5 TRNA
ST1-R5 mRNA (dynabeads)		ST1-R5 aRNA (RiboAmp from TRNA)
ST1-R5 mRNA (oligotex)	vs	ST1-R5 aRNA (RiboAmp from TRNA)
ST1-R5 mRNA (dynabeads)		ST1-R5 aRNA (RiboAmp from mRNA)
ST1-R5 mRNA (oligotex)		ST1-R5 aRNA (RiboAmp from mRNA)
ST1-R5 aRNA (RiboAmp from TRNA)		ST1-R5 aRNA (RiboAmp from mRNA)

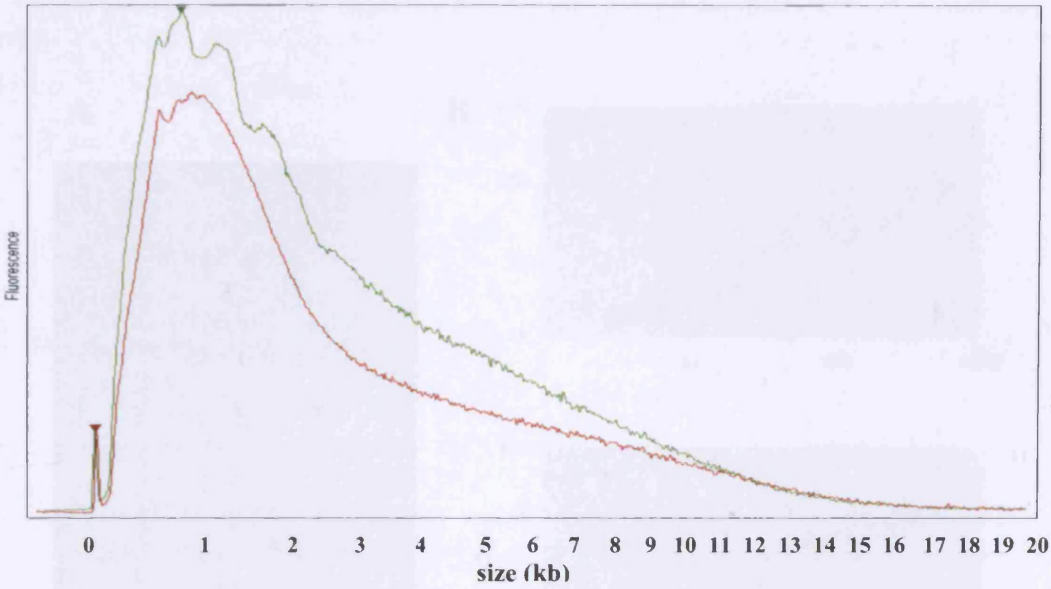
Figure 4.4 Evaluation of the process of RNA amplification and determination of optimal hybridisation conditions.
(A) Schematic of series of microarray hybridisations performed to determine optimal amplification protocol. **(B)** Graphs displaying SNR correlations for Cy5 and Cy3 channels, representing ST1-R5 TRNA vs. ST1-R5 TRNA; ST1-R5.R5 mRNA (dynabeads) vs. ST1-R5.R5 aRNA (RiboAmp from TRNA); ST1-R5.R5 mRNA (oligotex) vs. ST1-R5.R5 aRNA (RiboAmp from TRNA); ST1-R5.R5 mRNA (dynabeads) vs. ST1-R5.R5 aRNA (RiboAmp from mRNA); ST1-R5.R5 mRNA (oligotex) vs. ST1-R5.R5 aRNA (RiboAmp from mRNA); and ST1-R5 aRNA (RiboAmp from TRNA) vs. ST1-R5 aRNA (RiboAmp from mRNA).



For each array, the data was filtered by signal to noise ratio (SNR) in each channel. Ideally, there would be 100% correlation between TRNA and aRNA hybridisations and the overall correlation coefficient would be equal to 1. The scattering of values around the identity line reflects the influence of gene specific RNA amplification factors and experimental noise. Hybridisation of TRNA against itself gives a correlation coefficient of 0.93, with a tight distribution of values around the identity line (Figure 4.4 B-1). This indicates that the labelling and hybridisation of these samples was effective and does not introduce a significant bias into the data. Hybridisation of aRNA amplified from mRNA (Dynabeads) with aRNA amplified from TRNA (Picopure) also gives a high correlation coefficient (0.92). This implies that amplification from either mRNA extracted from TRNA, or straight from TRNA, does not adversely influence the resulting hybridisation profile (Figure 4.4 B-2). In all cases, for both channels, the scattering of values is equally distributed on both sides of the identity line, which indicates an overall linear amplification (Figure 4.4 B-3-6). Correlation coefficients are roughly equivalent, ranging from 0.74 to 0.83, regardless of reference RNA source and amplification substrate. Thus, there is a good correlation between signal intensities resulting from non-amplified mRNA as compared to aRNA, suggesting that aRNA hybridisations will correspond well, although not identically, to those obtained with unmodified samples. This is in agreement with the literature.

Bioanalyser analysis of amplified RNA indicate that the RiboAmp method also greatly improves the size profile of aRNA synthesised (compared to results with the Eberwine method, section 4.3.1.1, Figure 4.2-B). With RiboAmp, the majority of amplified transcripts were 300 - 400nt in length. Furthermore, RNA amplified directly from TRNA, as opposed to mRNA first isolated from TRNA (using Dynabeads), showed an improved size range, with an increased proportion of amplified transcripts being between 1000 and 5000 nt. This adverse effect of the Dynabead purification step is perhaps due to the preferential selection of high abundance short mRNAs by the Dynabeads prior to amplification. This would lead to an overall shortening of the RNA species represented in the amplified population. By amplifying from TRNA, however, all mRNA present is an amplification target and should therefore contribute to the amplified population (Figure 4.5).

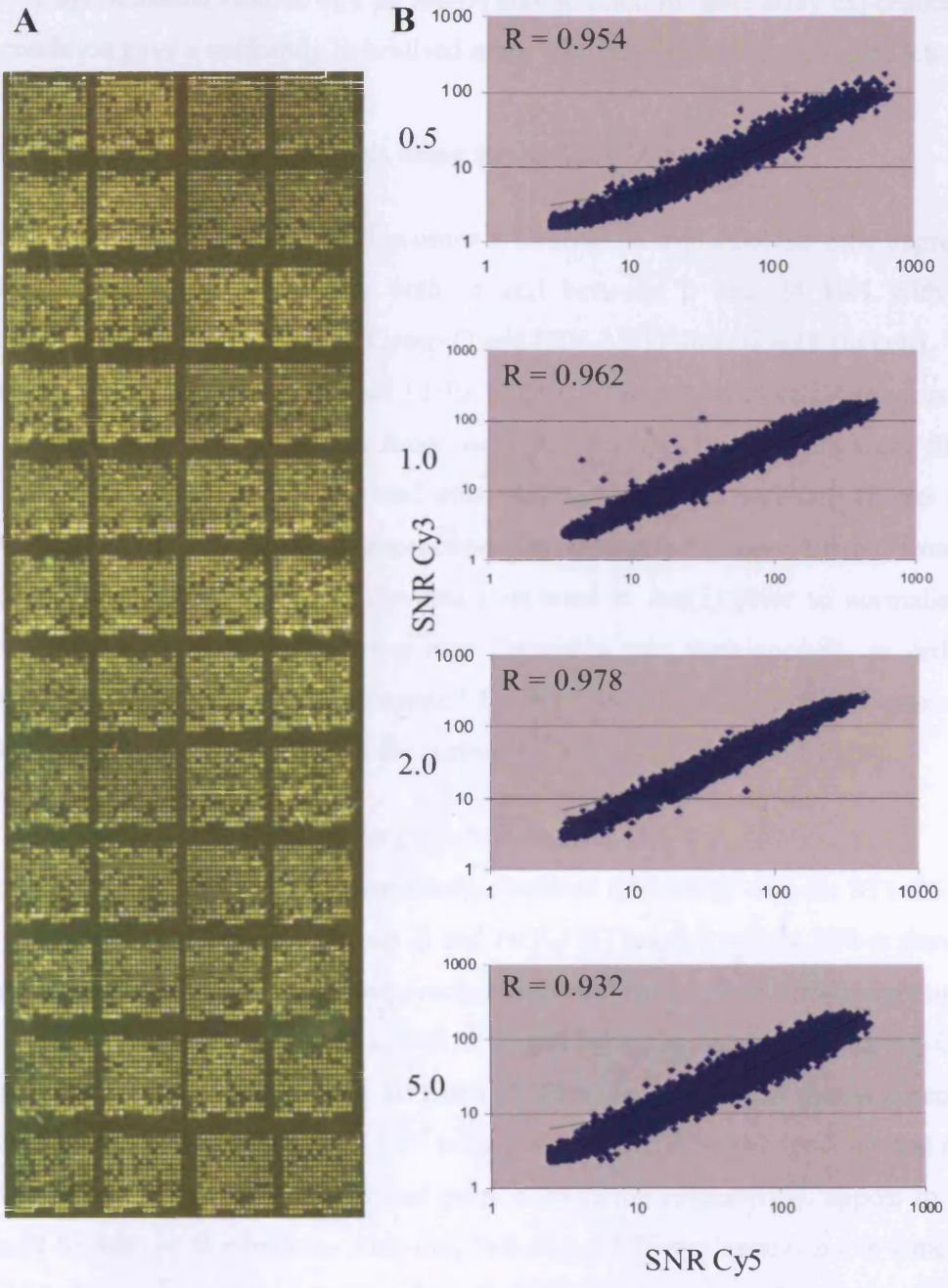
Figure 4.5 Comparison of the size profiles of amplified RNA, generated from either TRNA or mRNA.
Overlayed bioanalyser traces of aRNA produced using the RiboAmp method, from TRNA (extracted using Picopure RNA isolated kit - green) and mRNA (extracted from Trizol-extracted TRNA using Dynal dynabeads - red). The shoulder on the green trace (to the right) indicates an increased abundance of longer length aRNA transcripts in the RNA amplified from TRNA. This is reduced in the RNA amplified from mRNA.



4.3.1.4 Validation of reference RNA

Whilst mRNA was used as a control in the previous microarrays in order to assess the fidelity of amplification, for the arrays looking at the effect of HIV infection on host cell transcription an alternative reference was sought. This is because separate experiments become difficult to compare over time unless a common control, or reference RNA, of which the quality is highly controlled is used. A Universal Human Reference RNA (Stratagene, UHR-RNA) was selected, which is comprised of a collection of TRNA pooled from ten cell lines for optimal broad gene coverage. A recent report that described the use of aRNA in expression profiling analysis, however, outlined the importance of the test and reference samples being generated by an identical method: if using amplified RNA the reference must also be amplified (Li *et al.*, 2004b). Hence, a similar set of experiments were performed for the UHR-RNA as were previously used to establish fidelity of amplification. UHR-RNA was amplified using the RiboAmp system and hybridised to arrays with un-amplified UHR-RNA. Furthermore, in order to establish the optimal amount of reference to label and hybridise, a range of labelled UHR-aRNA concentrations were hybridised (5, 2, 1 and 0.5 μg). Ideally, if the UHR-RNA is optimal and amplification has not biased the mRNAs represented,

Figure 4.6 Analysis of the fidelity of RNA amplification of UHR-RNA.
(A) Composite array image of hybridisation of 1 µg UHR-RNA vs. amplified UHR-RNA. Uniform yellow colour indicates that amplification of the reference RNA is linear and that the broad hybridisation property of the TRNA reference has not been adversely affected by amplification. **(B)** Correlation coefficients for hybridisation of 0.5, 1, 2, and 5 µg UHR-RNA vs. amplified UHR-RNA. These data confirm the observation (in A) that the correlation between amplified and un-amplified reference is good.



the majority of the spots on the array should be yellow and give a strong and uniform signal (Figure 4.6). An example of such an array image is shown in Figure 4.6-A, where the majority of array elements are strongly co-hybridised. This indicates a broad and strong hybridisation pattern, which is preserved in aRNA. This is confirmed by analysis of the correlation coefficients between the signal to noise ratios for each channel (Figure 4.6-B). These correlations are consistently high, ranging from 0.93 to 0.98. A hybridisation amount of 1 μg aRNA was selected for later array experiments, as this condition gave a uniformly hybridised array with low background (Figure 4.6-A).

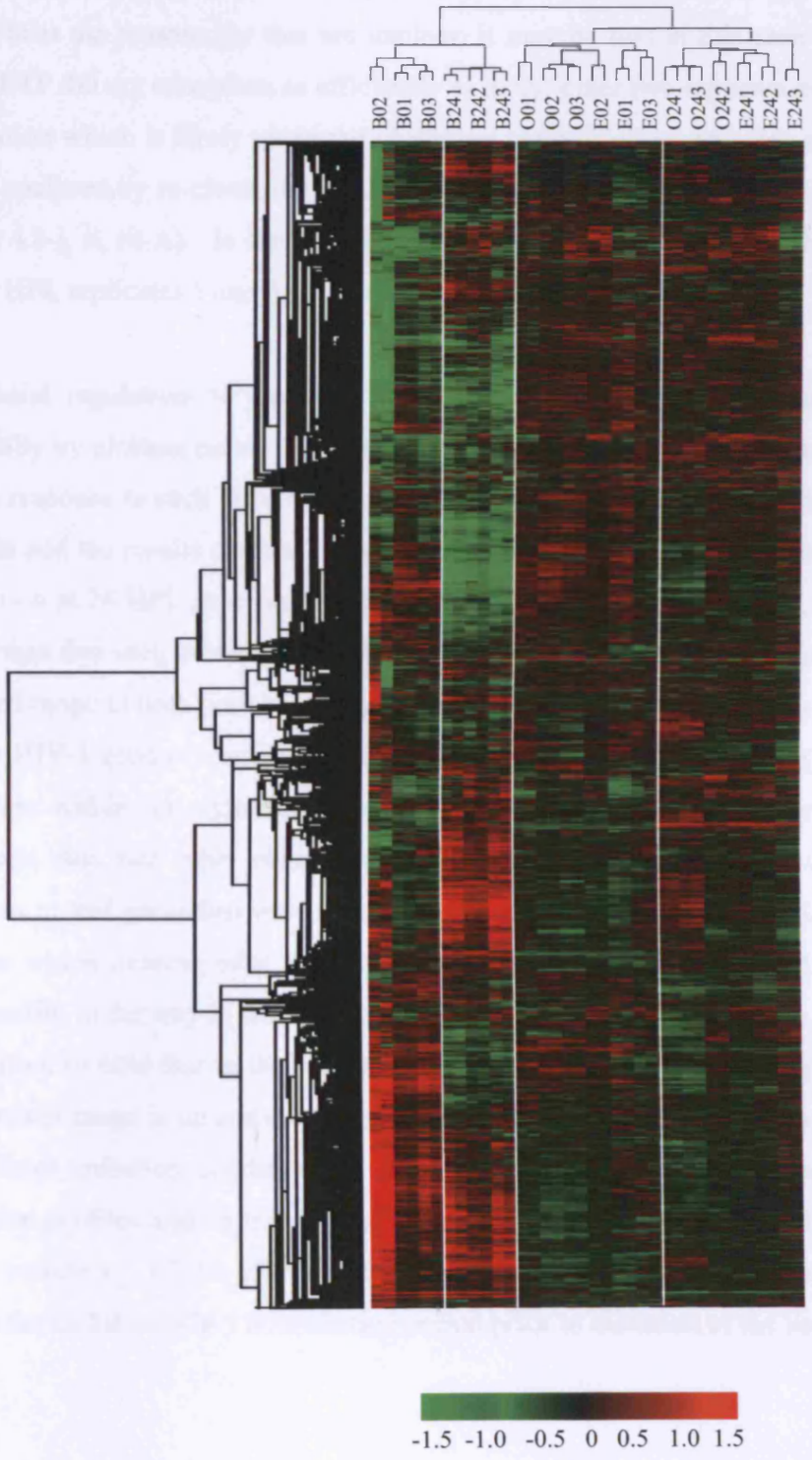
4.3.2 HIV gene expression analysis using spotted cDNA arrays

cDNA microarrays were performed in order to analyse, in triplicate, the gene expression profile of cultured ST1-R5 cells both at and between 0 and 24 HPI with high multiplicities of HIV-1 subtype B, Group O and HIV-2 ETP infection (1 ffu/cell). What is termed 0 HPI in this case is in fact 3 HPI, as cells were pre-incubated with virus for 3 h prior to washing and plating in fresh medium. The final array data were filtered according to signal to noise ratio and clustered using Cluster software (Eisen *et al*, 1998). Array elements for which expression measurements had been filtered from any of the arrays were removed and the data converted to log(2) prior to normalisation (median centring). A self-organising map algorithm was then applied, in order to control the orientation of nodes generated by average-linkage hierarchical clustering, and the results were visualised with the software Treeview (Eisen et al., 1998).

4.3.2.1 HIV induces clear changes in gene expression by 24 HPI

The complete dataset of filtered, normalised, clustered microarray data for ST1-R5 cells infected with HIV-1 subtype B, group O and HIV-1 ETP and 0 and 24 HPI is shown in Figure 4.7. The arrays cluster appropriately; duplicate datasets (per array) representing three replicates of each time point i.e. B01, B02 and B03, cluster together. Group O and HIV-2 (ETP) cluster together at 0 HPI and 24 HPI, but subtype B forms a separate branch within which 0 HPI and 24 HPI branches diverge. The strongest up and down regulated genes, represented by red and green colouration respectively, appear to be in response to subtype B infection. This may reflect the different conditions in which the initial 3 h virus pre-incubation was performed: the differences in viral titre meant that

Figure 4.7 Complete image of filtered, normalised, clustered microarray data for ST1-R5 cells infected with HIV-1 subtype B (B), group O (O) and HIV-1 ETP (E).
The dendogram on the left of the image represents genes clustered due to their similarity in expression level. The dendogram along the top of the image represents arrays clustered according to their similarity to other arrays. It can be seen that the arrays cluster appropriately in that the duplicate arrays representing three replicates of each time point i.e. B01, B02 and B03, cluster together. Group O and HIV-2 (ETP) cluster together at 0 HPI and 24 HPI, but subtype B forms a separate branch within which 0 HPI and 24 HPI branches diverge. The strongest up and down regulated genes, represented by red and green colouration, respectively, appear to be in response to subtype B infection.



the Group O and HIV-2 ETP infections were performed in almost neat infected cell supernatant, whereas in order to equalise volume the subtype B infection was performed in a higher concentration of fresh medium. Previous studies have shown that re-exposure to serum after starvation results in almost immediate changes in gene expression (Iyer *et al.*, 1999). It should also be noted that, although the tree structure shows the correct order in terms of grouping of viruses and time points, branch lengths occasionally show variation and complexity i.e. E243 (ETP 24 HPI, replicate 3), Figure 4.7. Whilst the reasons for this are unclear, it may be that in this case infection with HIV-2 ETP did not take place as efficiently as in the other two replicate experiments, an explanation which is likely given the clustering patterns observed when all arrays were further analysed by re-clustering the log(2) expression data individually, for each virus (Figure 4.8-i, ii, iii-A). In this representation of data, E243 clusters with E01 and E03 (ETP 0 HPI, replicates 1 and 3) (Figure 4.8-iii-A).

Differential regulation of genes at 24 HPI, compared to 0 HPI, is also presented graphically by plotting paired 0 and 24 HPI log(2) transformed expression data for each gene in response to each virus (Figure 4.8-i, ii, iii-B). This was ordered based on the 0 HPI data and the results confirm the observations in Figure 4.7: the response to subtype B infection at 24 HPI, in terms of magnitude of gene expression change, appears to be greater than that seen in response to HIV-1 Group O or HIV-2 infection, indicated by an increased range in both positive and negative log(2) expression changes (-1 to +1). The data for HIV-1 group O and HIV-2 ETP infection are similar in that log(2) expression values are within an equivalent range (-0.5 to +0.5). From this representation of expression data two other observations were made. Firstly, there is a trend in all infections in that genes that were most strongly up- or down regulated at 0 HPI seem to be those which become most strongly down- or up regulated at 24 HPI, indicating a commonality in the way in which the cells are responding to these viruses. Secondly, it is important to note that in the subtype B infected cells at 0 HPI there appears to be a much greater range in up and down regulated genes (Figure 4.8-i-B). This indicates that in the three infection conditions at 0 HPI the cells do not have equivalent gene expression profiles and in the case of subtype B gene expression levels show much greater variation. Whilst efforts were made to standardize the infection procedure, perhaps the inclusion of a 3 h incubation period prior to initiation of the time course, as

Figure 4.8 – i. Complete image of filtered, normalised, clustered microarray data for ST1-R5 cells infected with HIV-1 subtype B.
(A) Arrays cluster appropriately, with replicates of 0 HPI and 24 HPI forming distinct clusters. Clear clusters of up- and down-regulation of genes are apparent (marked by coloured blocks in the left dendogram). This data is illustrated in **(B)**, a graph representing differentially regulated genes at 24 HPI with subtype B HIV-1. Paired 0 and 24 HPI log2 transformed expression data for each gene was ordered based on the 0 HPI data. The grey histogram represents the range of gene expression values of infected cells at 0 HPI, and the blue bars represent how this is altered by 24 HPI.

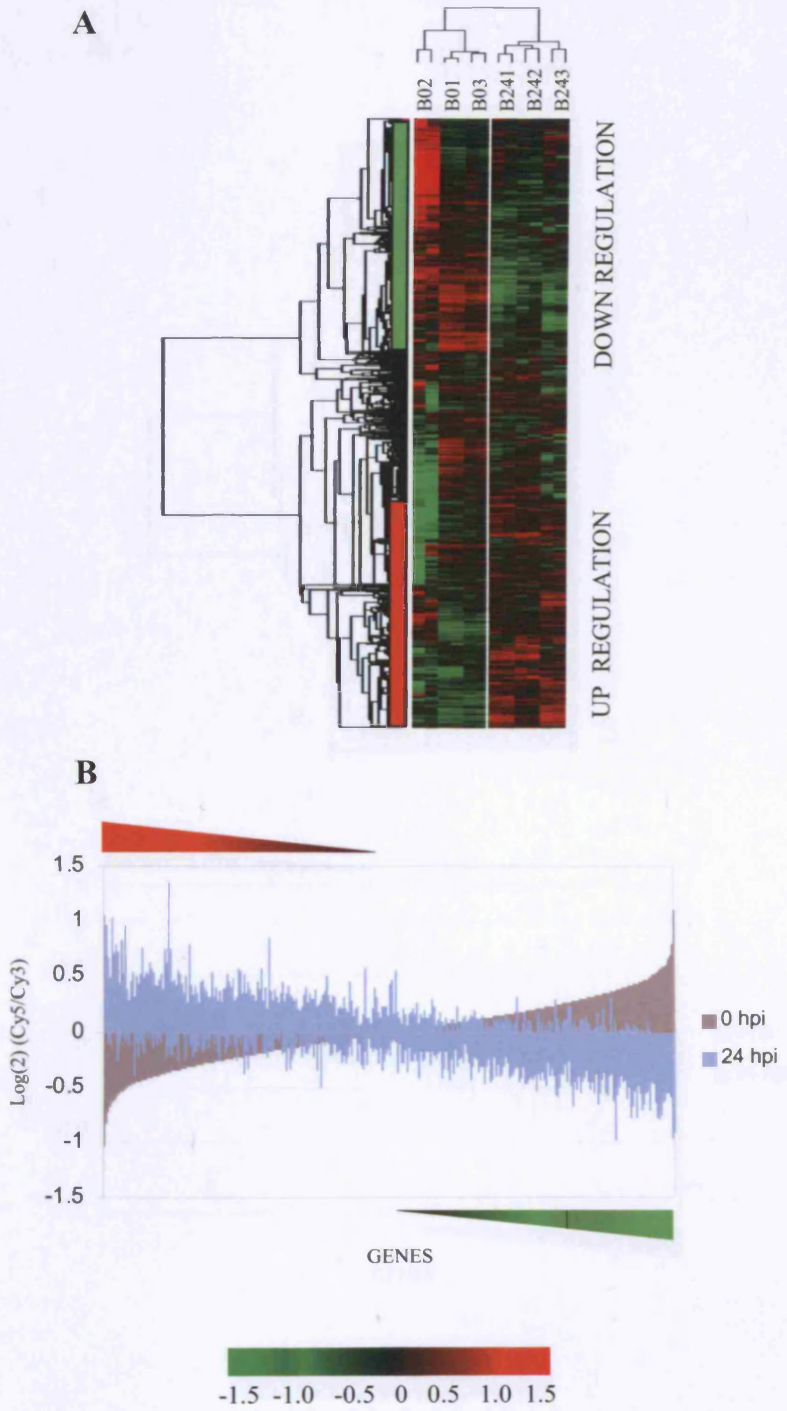


Figure 4.8 – ii. Complete image of filtered, normalised, clustered microarray data for ST1-R5 cells infected with HIV-1 Group O (O).
(A) Arrays cluster appropriately, with replicates of 0 HPI and 24 HPI forming distinct clusters. Clear clusters of up- and down-regulation of genes are apparent (marked by coloured blocks in the left dendogram). This data is illustrated in **(B),** a graph representing differentially regulated genes at 24 HPI with Group O HIV-1. Paired 0 and 24 HPI log2 transformed expression data for each gene was ordered based on the 0 HPI data. The grey histogram represents the range of gene expression values of infected cells at 0 HPI, and the blue bars represent how this is altered by 24 HPI.

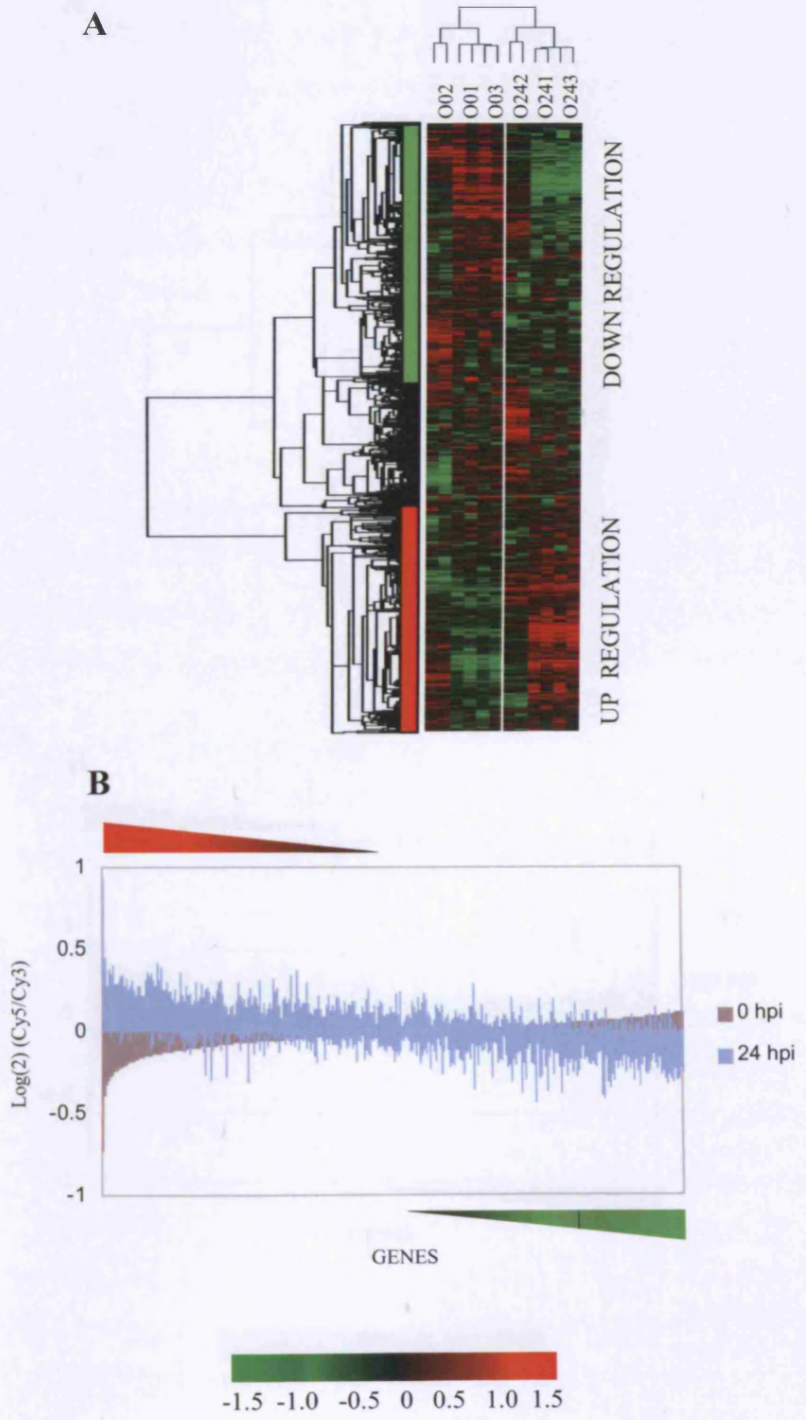
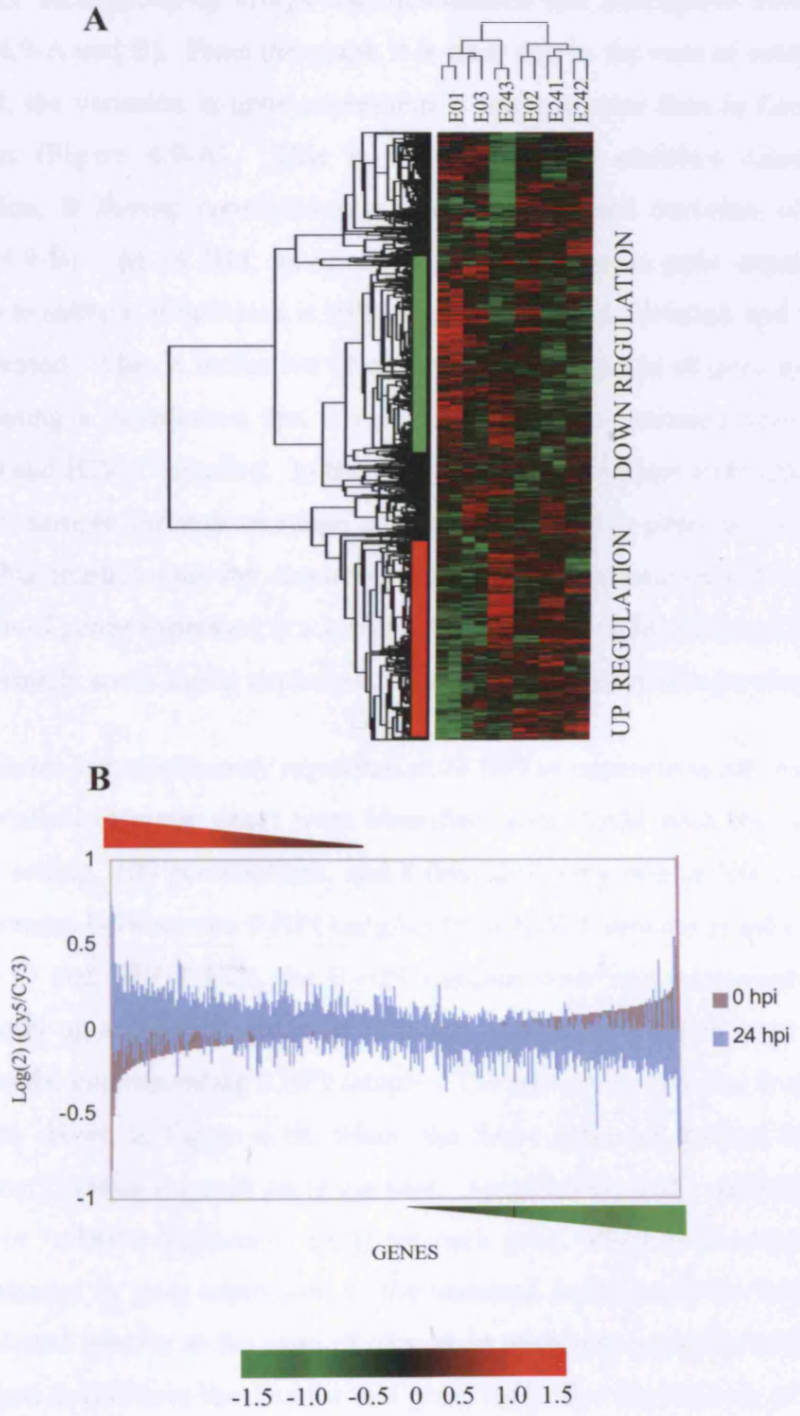


Figure 4.8 – iii. Complete image of filtered, normalised, clustered microarray data for ST1-R5 cells infected with HIV-2 ETP.
(A) Arrays cluster appropriately, with replicates of 0 HPI and 24 HPI forming distinct clusters. Clear clusters of up- and down-regulation of genes are apparent (marked by coloured blocks in the left dendogram). This data is illustrates in **(B),** a graph representing differentially regulated genes at 24 HPI with HIV-2 ETP. Paired 0 and 24 HPI log2 transformed expression data for each gene was ordered based on the 0 HPI data. The grey histogram represents the range of gene expression values of infected cells at 0 HPI, and the blue bars represent how this is altered by 24 HPI.



described, is responsible for this. This is important to bear in mind as later statistical analysis may be affected by such factors.

4.3.2.2 Gene expression at 0 HPI is also variable

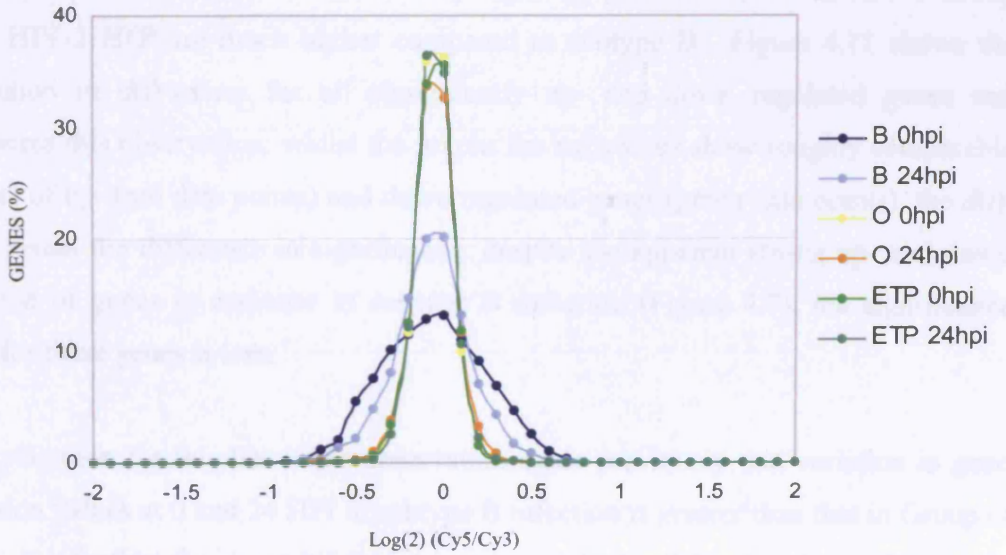
To further describe this variability in gene expression in response to the three different infecting viruses at both 0 HPI and 24 HPI, the distribution in mean log(2) expression values for each group of arrays was determined and descriptive statistics calculated (Figure 4.9-A and B). From the graph it is clear that in the case of subtype B infection, at 0 HPI, the variation in gene expression is much greater than in Group O or HIV-2 infections (Figure 4.9-A). This is reflected in the statistics calculated for each distribution, B having approximately twice the standard deviation of O and HIV-2 (Figure 4.9-B). At 24 HPI, however, whilst the range in gene expression values in response to subtype B infection is still large, the standard deviation and sample variance has decreased. This is indicative of an overall convergence of gene expression by 24 HPI, creating a distribution that is more similar to the variation seen in response to Group O and HIV-2 infection. In the case of these two viruses little change in standard deviation, sample variance or range of gene expression is observed between 0 and 24 HPI. This implies that the response to infection is characterised by a general up regulation of genes expressed at a low level prior to infection, balanced by a decrease in genes normally more highly expressed. Hence these data reflects previous observations.

4.3.2.3 Genes are significantly regulated at 24 HPI in response to all viruses

Differentially expressed genes were identified using SAM with the two-class paired response setting, 100 permutations, and a false discovery rate of less than 5%. Due to the differences between the 0 HPI samples from HIV-1 subtype B infection, compared to group O and HIV-2 ETP, the 0 HPI samples were not amalgamated - rather the significantly up and down regulated genes at 24 HPI were determined for each virus, relative to the corresponding 0 HPI sample. The pairwise, per virus, 0 vs. 24 HPI SAM results are shown in Figure 4.10, where the SAM plots are derived from analysis of filtered log(2) ratios for each array element. Specifically, SAM calculates an observed d-value, or "relative difference" ($d(i)$) for each gene, which is a statistic based on the ratio of change in gene expression to the standard deviation in the data for that gene. This is plotted relative to the ratio of change in randomly permuted gene expression to the standard deviation in the data for that gene, $d_E(i)$. For the majority of genes, $d(i)$

Figure 4.9 Evaluation of variation in gene expression data at 0 and 24 HPI.
(A) Graph representing the distribution of log(2) ratios for each array element (after filtering) for ST1-R5 cells infected with HIV-1 subtype B, Group O and HIV-2 ETP, at both 0 and 24 HPI. HIV-1 subtype B results in a larger distribution in genes both up and down regulated, in comparison to reference at 0 and 24 HPI. The distribution is smaller at 24 HPI. Cells infected with HIV-1 group O and HIV-2 ETP display a similar distribution of both up and down regulated genes, at 0 and 24 HPI. The plot indicates that there is no appreciable difference in distribution of gene expression values in cells infected with these two viruses at the two time points studied. **(B)** Table of general statistics derived from the log(2) expression data for each array element (after filtering) for ST1-R5 cells infected with HIV-1 subtype B, Group O and HIV-2 ETP, at both 0 and 24 HPI. The mean, standard deviation and sample variance and range of values are all greater for subtype B infection, than for O and ETP and both 0 and 24 HPI.

A



B

	B0	B24	O0	O24	ETP0	ETP24
Mean	-0.009	-0.011	-0.004	-0.006	-0.008	-0.004
Standard Deviation	0.293	0.223	0.108	0.123	0.113	0.106
Sample Variance	0.086	0.050	0.012	0.015	0.013	0.011
Range	2.150	2.358	1.619	1.776	1.869	1.437

$\approx d_E(i)$ i.e. there is no change in expression level between the two samples, but some genes are represented by points displaced from the $d(i) = d_E(i)$ line by a distance greater than a threshold, Δ, depicted by broken lines. This value was computed individually for each analysis and corresponds to a defined false discovery rate in this case of 5%, in the significantly regulated genes. These outliers therefore represent the significantly up and down regulated genes (Tusher *et al.*, 2001) (Figure 4.10).

This analysis shows that significantly regulated genes were found in response to all viruses at 24 HPI, and that infection with HIV-1 group O and HIV-2 ETP resulted in more significant up- and down regulation at 24 HPI than infection with HIV-1 subtype B, as $d(i)$ values calculated for significantly regulated genes 24 HPI with HIV-1 Group O and HIV-2 ETP are much higher compared to subtype B. Figure 4.11 shows the distribution in $d(i)$ values for all significantly up- and down regulated genes and emphasises this observation: whilst the graphs for each virus show roughly comparable numbers of up- (red data points) and down regulated genes (green data points), the $d(i)$ values reveal the difference in significance: despite the apparent strong up- and down regulation of genes in response to subtype B infection (Figure 4.7), the significance values for these genes is less.

The explanation for this lies if the observation made previously that variation in gene expression values at 0 and 24 HPI in subtype B infection is greater than that in Group O or HIV-2 infection (Figure 4.8/4.9). As mentioned, the $d(i)$ value for each gene is calculated based on the ratio of change in gene expression to standard deviation in the data. For subtype B the standard deviation of the gene expression distribution is greater due to the increased variation in expression values, so the calculated ratio of change in gene expression to standard deviation in the data is greater. Hence the $d(i)$ value for each gene is smaller; in effect (for subtype B) the significance of changes in gene expression are masked by the variation in the 0 HPI HIV-1 subtype B data. In interpreting the $d(i)$ value for each gene care must be taken, therefore, as if the same gene were similarly up regulated in response to subtype B and group O viruses by virtue of the fact that the standard deviation in the group O data is smaller than that in the subtype B data, the regulation of that gene would appear more significant in the case of infection with Group O. Consequently, in further analysis the $d(i)$ value for each gene will be taken as a reflection of both significance and variation within the data, but not an

Figure 4.10 Significance analysis of microarray (SAM) plots derived from analysis of filtered log(2) ratios for each array element, for ST1-R5 cells infected with HIV-1 subtype B, Group O and HIV-2 ETP.

The graphs show observed $d(i)$ and expected $d_E(i)$ values: the blue line indicates where $d(i) = d_E(i)$ and the dotted lines represent the cut-off Δ. Where genes are represented by points displaced from the $d(i) = d_E(i)$ line by a distance greater than a threshold, Δ, these are called significantly up regulated (in red) or down regulated (in green) at 24 HPI, as compared to 0 HPI. Δ was computed individually for each analysis so each allows for a false discovery rate of 5%, in the significantly regulated genes. These data show significantly regulated genes are found in all cases, and that infection with HIV-1 group O and HIV-2 ETP results in more significant up and down gene regulation at 24 HPI than infection with HIV-1 subtype B.

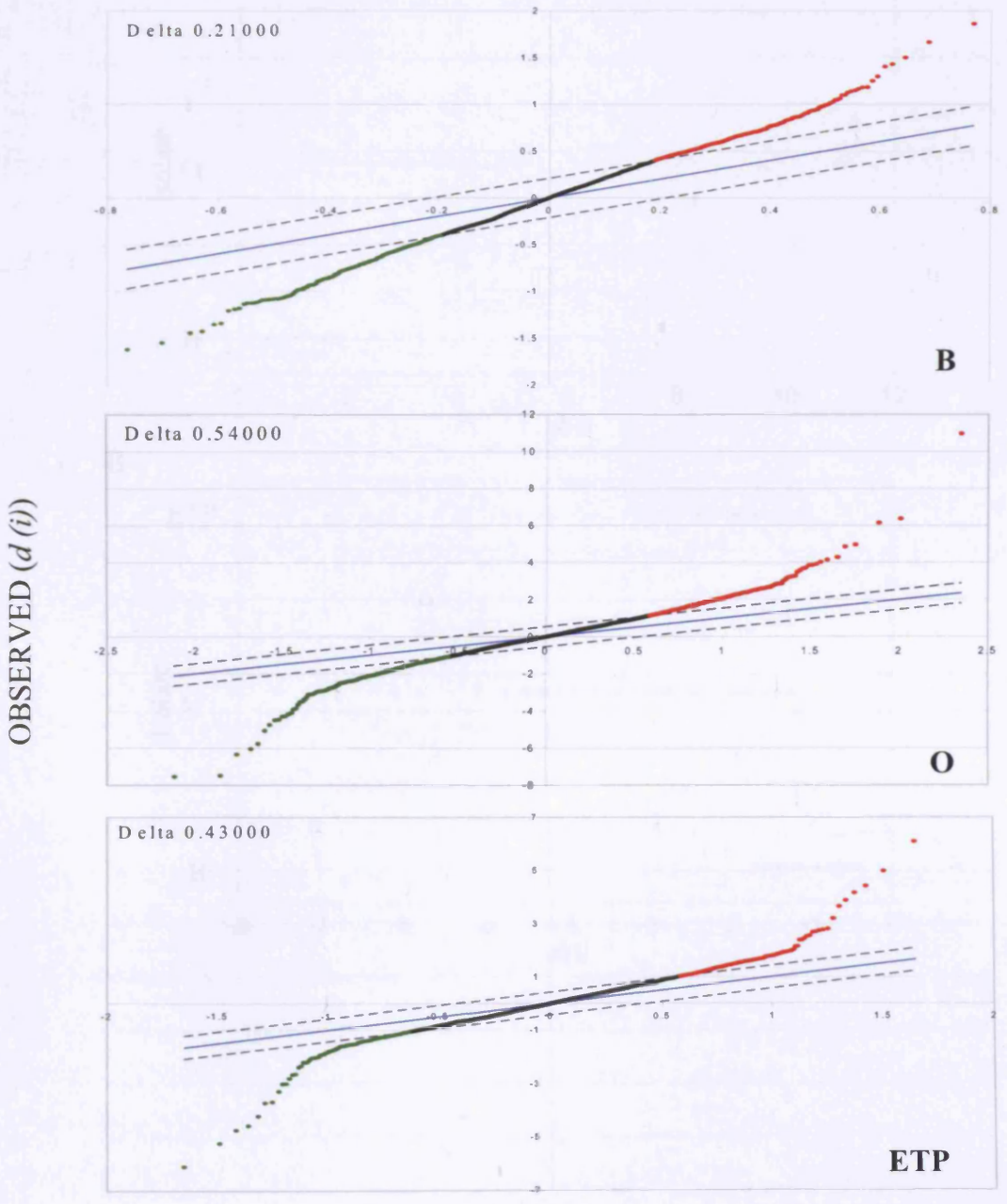
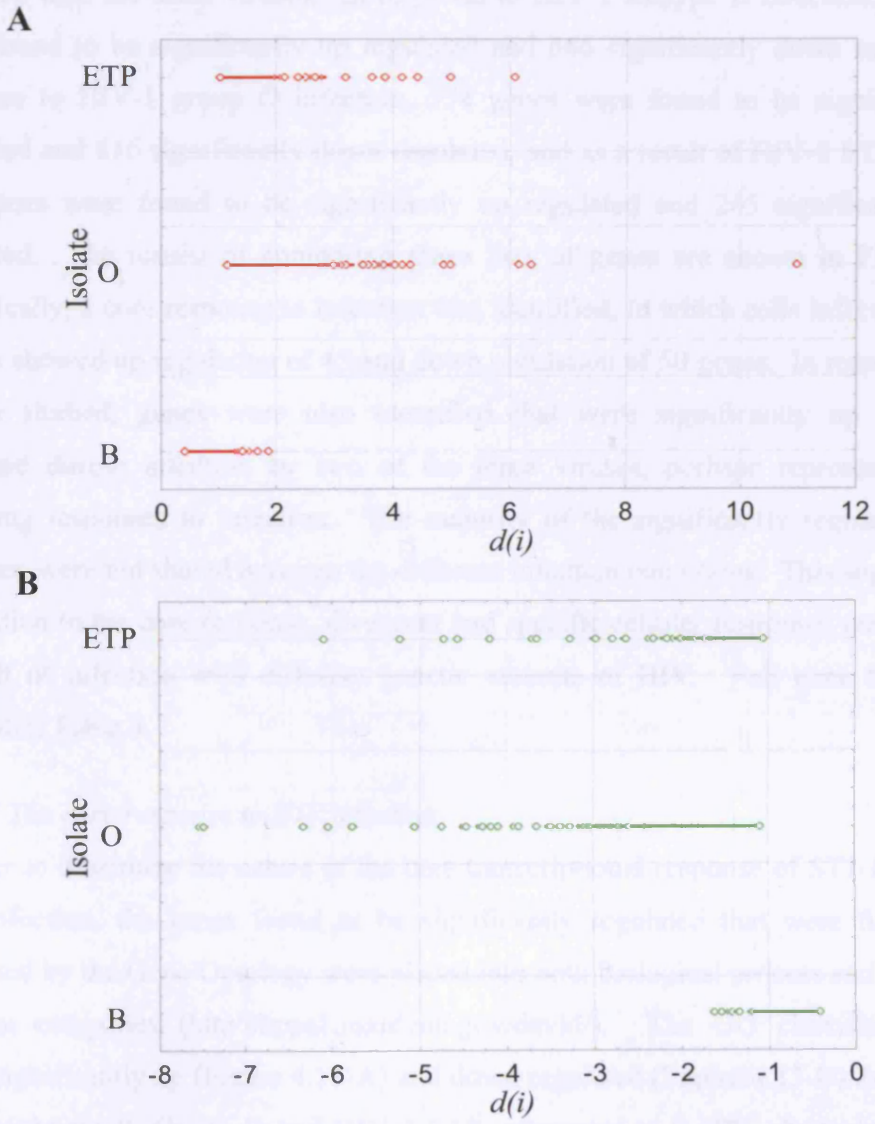


Figure 4.11 Distribution of *d*(*i*) values for differentially regulated genes in response to HIV-1 subtype B, group O and HIV-2.
Chart indicates the distribution of *d*(*i*) values for each gene generated by SAM for significantly up regulated **(A)** and down regulated **(B)** genes at 24 HPI, as compared to 0 HPI, in ST1-R5 cells infected with HIV-1 subtype B (B), group O (O) and HIV-2 (ETP).



absolute measure of significance. The numerical order of significant genes in terms of $d(i)$ value will be interpreted as the order of significance.

4.3.2.4 Evidence for core and diverging responses

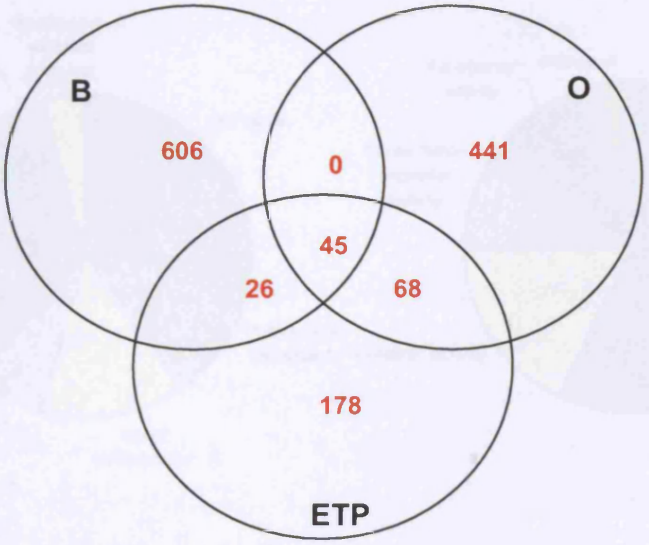
The lists of genes significantly up- and down regulated at 24 HPI were compared to determine the relationship between the transcriptional responses of ST1-R5 cells to infection with the three viruses. In response to HIV-1 subtype B infection, 677 genes were found to be significantly up regulated and 646 significantly down regulated. In response to HIV-1 group O infection, 554 genes were found to be significantly up regulated and 616 significantly down regulated, and as a result of HIV-2 ETP infection 317 genes were found to be significantly up regulated and 245 significantly down regulated. The results of comparing these lists of genes are shown in Figure 4.12. Specifically, a core response to infection was identified, in which cells infected with all viruses showed up regulation of 45 and down regulation of 50 genes. In response to the viruses studied, genes were also identified that were significantly up and down regulated during infection by two of the three viruses, perhaps representing more diverging responses to infection. The majority of the significantly regulated genes, however, were not shared between the different infection conditions. This suggests that, in addition to the core response, divergent and specific cellular responses may occur as a result of infection with different genetic variants of HIV. Full gene lists are in Appendix, Table 3.

4.3.2.5 The core response to HIV infection

In order to determine the nature of the core transcriptional response of ST1-R5 cells to HIV infection, the genes found to be significantly regulated that were functionally annotated by the Gene Ontology were placed into both biological process and molecular function categories (http://apps1.niaid.nih.gov/david/). The GO classifications for genes significantly up (Figure 4.13-A) and down regulated (Figure 4.13-B) 24 HPI with HIV-1 subtype B, Group O and HIV-2 ETP, compared to 0 HPI, showed a range of effects of infection on cellular function. For up regulated core response genes these include biological processes: cell cycle, nucleic acid metabolism, signal transduction, biosynthesis, lipid metabolism, response to stress, protein metabolism, cell-cell signalling, immune response and response to external stimulus. Of these the molecular functions are categorised as: binding, signal transducer activity, catalytic activity,

Figure 4.12 Core and diverging responses to HIV-1 subtype B, group O and HIV-2 infection at 24 HPI.
(A) Venn diagram showing numbers of significantly up regulated genes at 24 HPI, as compared to 0 HPI, in ST1-R5 cells infected with HIV-1 subtype B (B), group O (O) and HIV-2 (ETP). **(B)** Venn diagram showing numbers of significantly down regulated genes at 24 HPI, as compared to 0 HPI, in ST1-R5 cells infected with HIV-1 subtype B (B), group O (O) and HIV-2 (ETP).

A



B

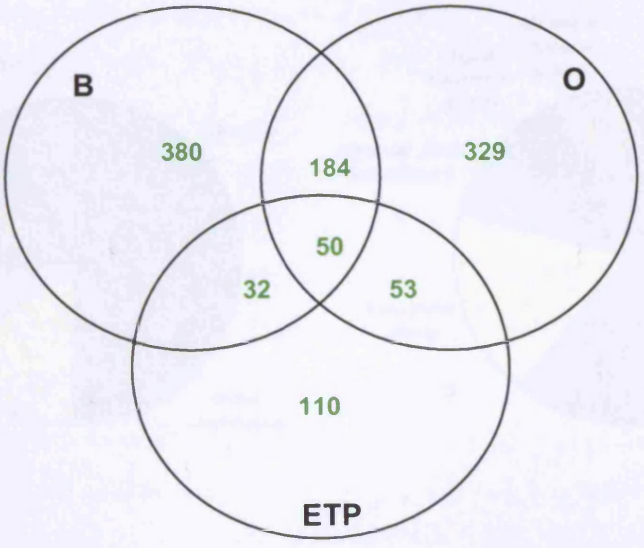
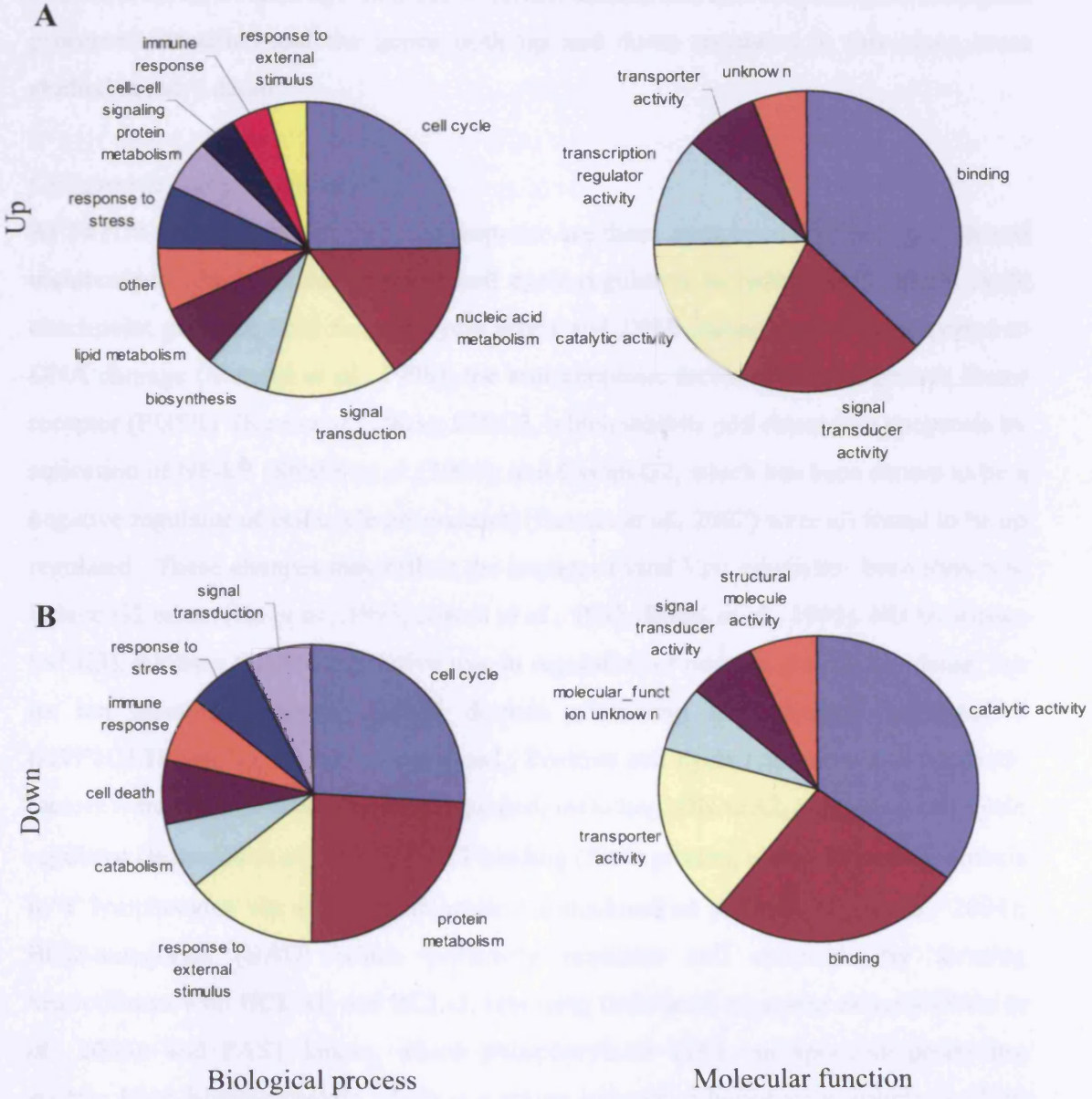


Figure 4.13 Classification of core response genes up- and down regulated in response to HIV-1 subtype B, Group O and HIV-2 ETP infection, 24 HPI.
Pie charts representing the Gene Ontology biological process and molecular function classifications for genes significantly up regulated **(A)** and down regulated **(B)** 24 HPI infection with HIV-1 subtype B, Group O and HIV-2 ETP, compared to 0 HPI. Only genes that have been annotated are included. These represent the 'core response'.



transcription regulator activity and transporter activity. Down regulated biological processes in this core response include: cell growth and/or maintenance, protein metabolism, response to external stimulus, catabolism, cell death, immune response, response to stress and signal transduction, and of these genes the molecular functions include: catalytic activity, binding, transporter activity, signal transducer activity and structural molecule activity. In order to further dissect this core response, the biological processes identified and the genes both up and down regulated in this class, were studied in more detail.

Cell growth and maintenance

At 24 HPI several genes in this core response are those associated with cell growth and maintenance. In particular negative cell cycle regulators including Rad1, a cell cycle checkpoint gene required for cell cycle arrest and DNA damage repair in response to DNA damage (Marathi *et al.,* 1998); the anti apoptotic factor epidermal growth factor receptor (EGFR) (Kari *et al.,* 2003); BIRC3, which inhibits p53 dependent apoptosis by activation of NF-kB (Stoffel *et al.*, 2004); and Cyclin G2, which has been shown to be a negative regulator of cell cycle progression (Bennin *et al.,* 2002) were all found to be up regulated. These changes may reflect the impact of viral Vpr, which has been shown to induce G2 arrest (He *et al.,* 1995, Jowett *et al.,* 1995, Rogel *et al.,* 1995). NIMA kinase (NEK3), a kinase that has a putative role in regulation of mitosis, was up regulated, but its ion channel substrate, FXYD domain containing ion transport regulator 1 (ATP1G1/HOMG2), was down regulated. Positive cell cycle regulators and apoptotic factors were correspondingly down regulated, including: DNAJA2, a positive cell cycle regulator (Edwards *et al.,* 1997); CD27-binding (Siva) protein, which induces apoptosis in T lymphocytes via a caspase-dependent mitochondrial pathway (Py *et al.,* 2004); BCl2-antagonist (BAD) which positively regulates cell apoptosis by forming heterodimers with BCL-xL and BCL-2, reversing their death repressor activity (Won *et al.,* 2003); and FAST kinase, which phosphorylates TIA1, an apoptosis-promoting nuclear RNA-binding protein which is a strong inducer of lymphocyte apoptosis (Tian *et al.,* 1995, Li *et al.,* 2004c). Tat-interacting protein 30 (TIP30), which is a novel serine/threonine kinase that phosphorylates the C-terminal domain of the largest RNA polymerase II subunit and induces the expression of apoptosis related genes Bad and Siva, was down regulated, a finding which corresponds with the observed down

regulation of BAD and Siva described. As TIP30 also interacts with amino acids 1-48 of HIV-1 Tat, including the Tat activation domain, and enhances Tat activation of the HIV-1 LTR promoter (Xiao *et al.,* 1998), its down regulation as part of this core response suggests an anti-apoptotic, or anti-viral mechanism not previously observed.

H-ferritin function is essential for the control of cell proliferation and transformation by c-MYC, a proto-oncogene transcription factor that can both activate and repress the expression of target genes. Specifically, c-MYC represses the expression of the heavy subunit of the protein ferritin (H-ferritin) which sequesters intracellular iron, and stimulates the expression of the iron regulatory protein-2 (IRP2) which increases the intracellular iron pool. Down regulation of H-ferritin during infection suggests that this effect of cell cycle regulation may also be important in HIV infection (Ameglio *et al.*, 1993, Wu *et al.*, 1999). In addition, annexin A2 is down regulated. In complex with S100A10, annexin A2 controls the distribution of transferrin receptor-containing recycling endosomes. Together these data allude to a relationship between HIV infection and iron metabolism (Zobiack *et al.,* 2003), but this is as yet poorly understood. It has been suggested that high iron status, both in blood and intracellularly (i.e. within macrophage) may adversely influence the outcome of HIV-1 infection and thus these data might suggest a direct role of HIV infection in this pathology (Gordeuk *et al.*, 2001).

Transcription, transcriptional regulation and translation

Genes involved in transcription, transcriptional regulation and translation were also inhibited during infection with all viruses, including the transcription factors p38 interacting protein and MondoA; DEAD box polypeptide 49 (an ATP-dependent RNA helicase); Pirin, an iron-binding nuclear protein and transcription cofactor (Wendler *et al.,* 1997); Peroxiredoxin 2, which regulates the signal transduction pathways that utilise AP-1 to influence cell growth and apoptosis (Butterfield *et al.,* 1999); and exportin, a protein belonging to the RAN-GTPase exportin family that mediates export of tRNA from the nucleus to the cytoplasm (Arts *et al.,* 1998). This general repression, in combination with the induction of cell cycle repressors and inhibition of cell cycle activators in response to HIV infection, agrees with previous observations (van t'Wout *et al.,* 2003). Interestingly, peroxiredoxin activity may also be capable of inhibiting HIV infection (Butterfield *et al.,* 1999).

Certain transcription and splicing factors, however, were up regulated implicating them in facilitating HIV production. These include: the transcription factor AP-2 gamma (activating enhancer binding protein 2 gamma), a sequence-specific DNA-binding transcription factor involved in the activation of several developmental genes, and c-fos, which encodes a leucine zipper protein that can dimerise with proteins of the JUN family, thereby forming the transcription factor complex AP-1; SERTA domain containing 2 (SERTAD2); and zinc finger protein 131. Interestingly, exposure to extracellular Tat has been shown to induce the expression of both c-fos mRNA and protein in serum-starved, non-infected Jurkat CD4+ lymphoblastoid T cells. This Tat-mediated induction of c-fos may therefore contribute to the immune hyper-activation that characterises the progression to AIDS, which constitutes the optimal environment for HIV-1 replication (Gibellini *et al.,* 2001). RNA binding motif protein 25 (S164) and splicing factor arginine/serine-rich 5 (SFRS5), which play roles in constitutive splicing and can modulate the selection of alternative splice sites, were also induced. In relation to nucleic acid biosynthesis, IMP (inosine monophosphate) dehydrogenase 1, the rate-limiting enzyme of *de novo* GTP biosynthesis (Natsumeda *et al.,* 1990) was also up regulated.

An increase in expression of Importin β, implicated by some studies in nuclear translocation of the pre-integration complex (PIC) allowing HIV to infect non-dividing cells, was detected (Gallay *et al.,* 1997, Depienne *et al.,* 2001). In addition several of the thyroid hormone receptor interacting protein (TRIP) family were up regulated. This is interesting, as Trip-1 has been shown to simultaneously modulate responses involving both cytokine and nuclear receptors in cells of erythroid lineage (Ingley *et al.,* 2001). Thyroid hormone receptor (TR), for example, binds to the integrated HIV-LTR in chromatin *in vivo* and represses the LTR in the absence of thyroid hormone (TH) by recruiting co-repressor complexes containing histone deacetylases. Upon TH binding, TR causes chromatin remodelling and LTR activation, probably in cooperation with other cofactors such as Tripl (Ishizuka *et al.,* 2001). This up regulation of TH-responsive factors may therefore represent a mechanism during HIV lytic replication to ensure HIV-LTR activity (Hsia and Shi, 2002).

Response to external stimulus and cell signalling

In response to infection the general level of gene expression related to cell signalling and activation, in response to stress, was increased. The growth factors vascular endothelial growth factor (VEGF) and VEGF-C were up regulated, most likely induced by the action of HIV Tat, which has been demonstrated *in vitro* (Benelli *et al.*, 2000). Growth factor receptor-bound protein 10 (Grb10), which acts as a positive regulator in VEGF-R2 signalling and protects VEGF-R2 from degradation by interacting with Nedd4, a component of the endocytic machinery, was also up regulated (Murdaca *et al.*, 2004). Coincident with this up regulation, possibly through Tat, an up regulation in tissue inhibitor of matrix metalloproteinase 2 (TIMP2), a peptidase involved in degradation of the extracellular matrix, was observed. On one hand, this factor has been shown to partially inhibit the monocyte invasion induced as a result of VEGF expression, perhaps suggesting some form of host cell response to Tat induced gene expression changes (Lafrenie *et al.*, 1996). However, TIMP-2 also forms a receptor in complex with the membrane-type-1-matrix-metalloproteinase (MT1-MMP). This regulates the generation of functionally active matrix metalloproteinase 2 (MMP-2) (Murphy *et al.*, 1999). It is known that combined Tat and fibroblast growth factor 2 (FGF2) augment active MMP-2 release by increasing the levels of both activated MMP and cell membrane-associated TIMP-2, whilst decreasing the amount of secreted TIMP-1 and -2. These *in vitro* effects are associated with the induction of vascular permeability and oedema *in vivo* by combined Tat and FGF2, and are thought to be key factors in the vasculopathy of HIV-1-infected individuals (Toschi *et al.*, 2001).

Immune response

Receptors and signalling molecules involved in immune responses are modulated during HIV infection. As part of this core HIV transcriptional response, down regulation of major histocompatibility complex (MHC), class I-C was detected, which has been directly attributed to the affect of HIV Vpu (Kerkau *et al.*, 1997) and Nef (at the protein level - Andrieu *et al.*, 2001). The activity of Nef in down regulating MHC, however, requires peroxisomal acyl-CoA thioesterase (hTE), which binds to a region of the Nef core in close proximity to the putative CD4 binding site. Nef mutants disrupted in this binding domain were found to be defective in MHC class I down modulation and enhancement of viral infectivity (Liu *et al.*, 2000, Cohen *et al.*, 2000b). Down regulation of h(TE) in response to all viruses was detected, therefore suggesting that Nef

exploits a cellular pathway for MHC-1 surface expression that can be regulated transcriptionally by the removal of a necessary cofactor. Not all Nef alleles bind to hTE with high affinity, however, so the role of hTE during HIV infection in far from clear (Cohen *et al.*, 2000b). Nef has also been shown to induce CD4 down regulation (Greenberg *et al.*, 1997), and whilst CD4 transcript down regulation was not detected in this core response, the adaptor-related protein complex 2 (AP2) which belongs to the adaptor complexes medium subunits family, was up regulated. It has previously been shown that Subunit H of the V-ATPase binds to the medium chain of AP2 and connects Nef to the endocytic machinery, allowing it to carry CD4 from clathrin coated pits at the cell surface to lysosomes, for its degradation (Geyer, 2002). Up regulation of AP2 may therefore be part of the viral modulation of endocytic processes that, amongst other things, results in down regulation of CD4. Together these data suggest that HIV protein-mediated down regulation of MHC-1 and CD4 at the cell surface are augmented by transcriptional modulation of cellular pathways that are part of the same process.

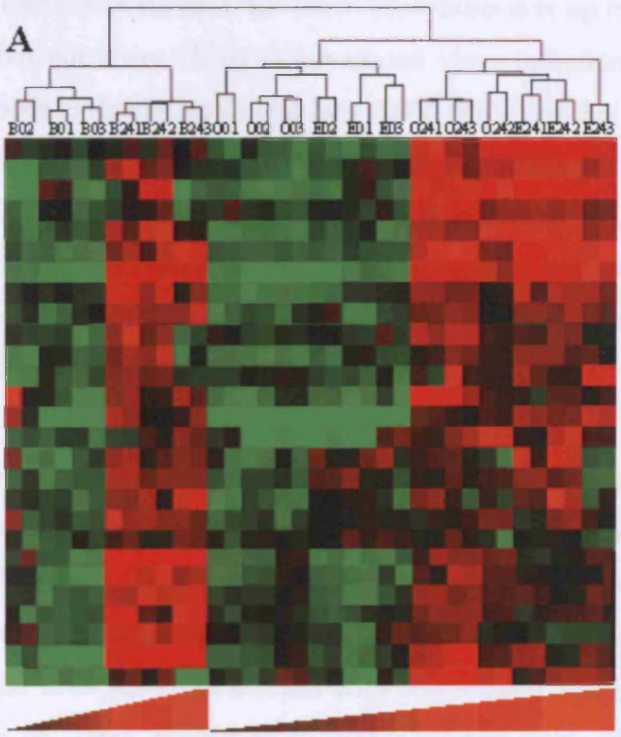
In addition, AP2 is important in the process of gp120 internalisation from the cell membrane, shortly after its delivery there from the endoplasmic reticulum. It is thought that this rapid internalisation of envelope proteins may be disadvantageous to the virus in that it prevents viral assembly at the cell membrane. It may be a function of Gag binding to the cytoplasmic domain of gp41, therefore, which prevents its recognition by the intracellular transport machinery and hence its internalisation, allowing viral assembly and budding at the cell membrane. An up regulation of AP2 may therefore be a mechanism by which the host can overcome this 'masking' effect and enable internalisation of gp120, disrupting viral budding. Conversely, this up regulation of AP2 may be advantageous to the virus, as internalisation of envelope by AP2 may result in its re-direction to a sub-cellular compartment where HIV budding also takes place (Morita and Sundquist, 2004).

Lipid and protein metabolism and intracellular trafficking

One of the most striking patterns of gene modulation directly resulting from infection of ST1-R5 cells by all viruses studied, was the up regulation of genes encoding enzymes involved in the cholesterol biosynthesis pathway. These genes are among the most strongly regulated elements detected on the arrays, this cluster shown in Figure 4.14 including; lysophosphatidic acid acyltransferase-delta; 24-dehydrocholesterol reductase;

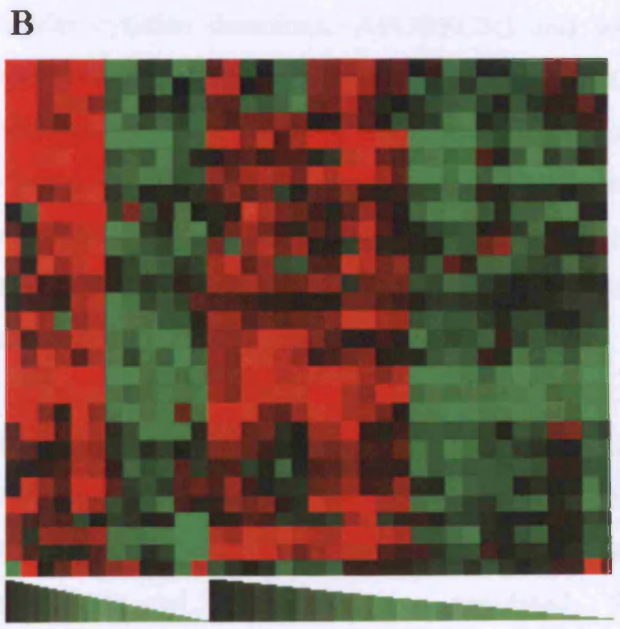
acetyl-Coenzyme A acetyltransferase 2 (acetoacetyl Coenzyme A thiolase); 3-hydroxy-3-methylglutaryl-Coenzyme A reductase (3-HMG-CoA reductase); and mevalonate kinase. It is well established that the HIV envelope has a higher cholesterol/phospholipid ratio than the host cell plasma membrane and that removing cholesterol from virions (Liao *et al.,* 2001, Ono & Freed, 2001) or inhibition of cholesterol biosynthesis with Lovastatin inhibits HIV-1 production both *in vitro* (Maziere *et al.*, 1994) and *in vivo* (del Real *et al.,* 2004). This statin inhibition is thought to be mediated via several mechanisms, including: disruption of cholesterol-rich domains on the cell surface (lipid rafts) which are proposed to provide platforms for specific regulated protein-protein interactions required for the process of viral assembly and through which HIV-1 buds during viral maturation (van 't Wout *et al.,* 2003); inhibition of actin cytoskeletal rearrangement by down regulation of Rho activity (del Real *et al.,* 2004); and by diminishing HIV-1 attachment to target cells by suppression of ICAM-1-LFA-1 interactions (Giguere and Tremblay, 2004). Since all cholesterol is provided by the host, it is reasonable to hypothesise that HIV may facilitate up regulation of intracellular cholesterol. Other microarray studies have demonstrated that HIV infection up regulates both the LDL receptor and many genes in the cholesterol biosynthesis pathway (van t'Wout *et al.,* 2003) and more specifically, the viral protein Nef has been shown to be a driving force in this up regulation of cholesterol biosynthesis (Zheng *et al.,* 2003). It is well known, however, that cytokines induce marked changes in lipid metabolism that lead to hyperlipidemia, which represents part of the innate immune response and may be beneficial to the host (Feingold *et al.,* 1998). Changes in lipid and lipoprotein metabolism may be beneficial in a number of ways including: lipoproteins competing with viruses for cellular receptors; apolipoproteins neutralizing viruses; lipoproteins binding and targeting viruses for destruction; redistribution of nutrients to cells involved in the immune response and/or tissue repair; and lipoproteins binding toxic agents and neutralizing their harmful effects. In the context of autoimmunity, it has been shown that inhibition of cholesterol biosynthesis using statins reduces the inducible expression of MHC class II molecules and blocks the expression of costimulatory molecules necessary for the induction of autoaggressive T cells (Youssef *et al.,* 2002). They also shift the balance of cytokines produced by autoaggressive T cells from T helper-1 (T_H1) proinflammatory cytokines, such as IFNγ and TNF, to T_H2-type cytokines, including IL-4, IL-5, IL-10, and IL-13 (Steinman *et al.,* 2004).

Figure 4.14 Cluster diagram showing subset of core response genes.
Cluster of genes found using SAM to be significantly up regulated **(A)** and down regulated **(B)** at 24 HPI with HIV-1 subtype B, Group O and HIV-2 ETP, compared to 0 HPI. These represent the 'core response' genes which include examples of the top 20 most significantly up and down regulated genes, by rank order of $d(i)$ value.



7-dehydrocholesterol reductase
mevalonate kinase
vascular endothelial growth factor
CD8 antigen, beta polypeptide 1
adaptor-related protein complex 2
acetyl-coenyme A acetyltransferase 2
3-hydroxy-3-methylglutaryl-coenzyme A
oncogene TC1
steroid sulphatase (microsomal)
KIAA0301
v-fos
zinc finger protein 220
KIAA0068
vascular endothelial growth factor
IMP dehydrogenase 1
Zinc finger protein 239
Serine/cysteine proteinase inhibitor
cDNA DRFZp586J021
RAD1
interferon inducible protein kinase
splicing factor, arginine/serine-rich 5
splicing factor, arginine/serine-rich 5
calmodulin 2

cyclin G2
transcription factor AP2 gamma
tissue inhibitor of metalloproteinase



HLA-B associated transcript-3
FK506 binding protein-4
trophinin associated protein
defender against cell death 1
CD27-binding protein (SIVA)
peroxiredoxin 5
ferritin
cell division cycle 37 homolog
BCL-2 antagonist/killer 1
KiSS-1 metastasis suppressor
BE549032 similar to Hs.24054
leupaxin
thyroid hormone receptor interactor 4
DNA polymerase, beta subunit
vimentin
myosin, light polypeptide 6
proteasome (prosome, macropain) subunit
KIAA0368 protein
B cell RAG associated protein
annexin A2
cell cycle progression 2 protein
proteasome (prosome, macropain) subunit
gamma tubulin ring complex protein
syntaxin 4A
HIV-1 Tat interactive protein
peptidylprolyl isomerase (cyclophillin)
pirin
protein tyrosine phosphatase
FXYD domain-containing ion transport channel

Accordingly, up regulation of cholesterol biosynthesis early during viral infection may be important in the expansion of T_H1 proinflammatory responses – crucial in initial attempts to control viral replication. Recent collaborative efforts to understand this process have involved microarray studies of several different viral infections, *in vitro*. These data suggest that sterol biosynthesis is up regulated in lytic (HIV, HCMV, RSV), but not latent (KSHV) enveloped virus infections, and that this regulation can occur both early (HCMV) and late (HIV, RSV) in the virus life cycle. Regulation may also involve the entire biosynthetic pathway (HCMV, HIV) or specific genes (RSV) (van 't Wout *et al.*, 2004). It seems therefore, that up regulation of cholesterol biosynthesis during viral infection may be advantageous for both host and pathogen. Perhaps the co-evolution of both has resulted in this intimate relationship, the cause and effect of which will be very difficult to distinguish.

The proteasome subunit β6 precursor and proteasome subunit β6 itself, which cleaves peptides in an ATP/ubiquitin-dependent process in a non-lysosomal pathway, were both down regulated. As well as normal cellular protein degradation the proteasome is responsible for degradation of HIV integrase (Mulder and Muesing, 2000), and HIV Tat has been shown to interact with this subunit and interfere with its processivity (Seeger *et al.*, 1997, Huang *et al.*, 2002, Apcher *et al.*, 2003). Furthermore, Vif binds to the cellular cytidine deaminase APOBEC3G and targets it for degradation through an interaction with the proteasome (Sheehy *et al.*, 2003). Down regulation of this subunit therefore may be a result of Tat, and may promote viral infectivity by reducing the degradation rate of its viral proteins, such as integrase. Conversely, down regulation of the proteasome may reduce viral infectivity by circumventing the activity of Vif in targeting of the host anti-viral protein APOBEC3G for degradation, allowing it to be incorporated into virions and disable viral progeny.

Finally, modulation of intracellular trafficking occurs during this core response to HIV infection. In addition to AP2 being up regulated, potentially contributing to down regulation of surface CD4 expression by interaction with HIV Nef; SEC6-like 1, syntaxin 5a and Vam6 were down regulated. SEC6-like 1 is a component of the exocyst complex; a multiple protein complex essential for targeting exocytic vesicles to specific docking sites on the plasma membrane (Matern *et al.*, 2001). Vesicle docking is thought to be regulated, in part, by the specific interactions of syntaxin with a vesicle-

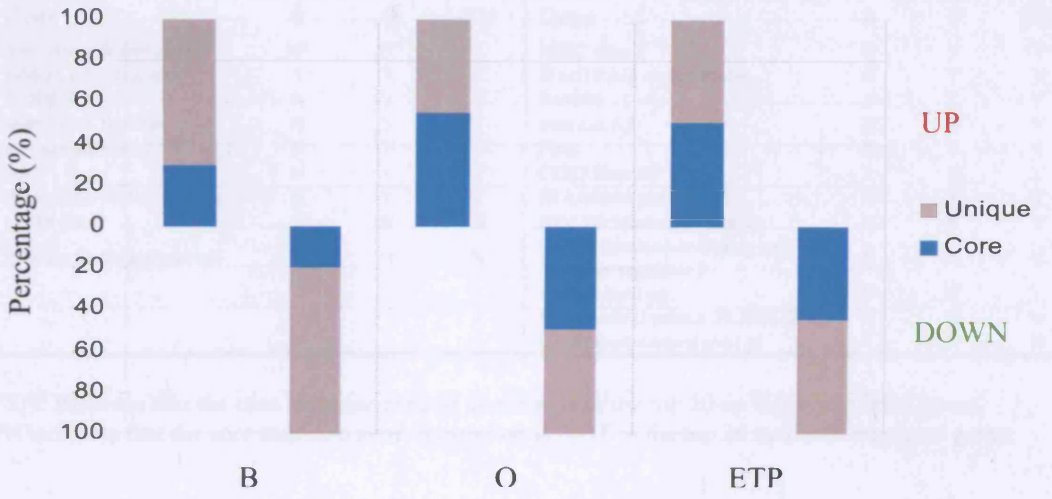
associated membrane protein, synaptobrevin/VAMP. Vam6 promotes lysosome clustering and fusion *in vivo* (Caplan *et al.*, 2001). Whilst these proteins have not previously been identified as being regulated in infection, there may be some functional relevance of this apparent down modulation of factors involved in vesicle docking and fusion. A further gene, Surfeit 4, was up regulated. In eukaryotic cells protein transport between the endoplasmic reticulum and Golgi compartments is mediated in part by non-clathrin-coated vesicular coat proteins (COPs) and whilst the specific function of Surfeit 4 has not been determined, its yeast homolog is directly required for packaging the soluble secretory protein glycosylated pro-alpha-factor into COPII vesicles (Belden and Barlowe, 2001).

Several genes were regulated as part of this core response to HIV infection, whose functions are unknown. These are genes with accession numbers: R11659 (up regulated), and; BE548179, R24842, BE547709 (down regulated). It will be interesting to see if, as these genes are characterised, they fall into place with any of the observations described or whether they will indicate other processes of host cell modulation either by HIV itself, or as part of this core response to HIV infection.

4.3.2.6 The 'significance' of the core response

The core response described, comprised of 95 both up and down regulated genes, may be considered to encompass both the fundamental responses of a host cell to HIV infection, and the more conserved ways in which HIV can modulate the gene expression of its host cell. It could be presumed therefore, that these most conserved host – pathogen interactions occurring during HIV infection may be associated with more consistent and strong examples of gene modulation, in experimental repeats. These examples would therefore be the most statistically significant, with high $d(i)$ values (section 4.3.2.3). In order to determine whether this is the case, the top twenty most significantly up and down regulated genes, as classified by SAM in response to HIV-1 subtype B, group O and HIV-2 ETP infection at 24 HPI, were identified. The number of these genes that are part of the core response was calculated and the results are shown in Figure 4.15.

Figure 4.15 The 'significance' of core response genes.
Graph shows the relative proportions of core response and unique genes (only regulated in response to one virus) in the top 20 most significantly regulated genes, as a result of infection.



Clearly, whilst core response genes are among the top 20 up and down regulated genes in response to B, O and ETP infection, they do not account for them all. In subtype B infection, core response genes comprise only 30% of the most significantly up- and 20% of the most significantly down regulated genes. In response to Group O and HIV-2 ETP infection, approximately 50% of the top 20 up and down regulated genes are those found to be part of the core response. It would seem that in addition to conserved and strong interactions between the host cell and HIV, there are significant processes of gene modulation occurring that are unique to each genetic variant of HIV. By examining these genes, potential differences between the ways in which HIV-1 subtype B, group O and ETP interact with the host cell may begin to characterised.

Prior to this analysis, however, these core response genes that comprise approximately 50% of the top 20 significantly up and down regulated genes in response to infection were analysed. The highly significant core response genes shown in Table 4.3 mainly include up regulated genes that are associated with lipid metabolism, cell cycle and intracellular trafficking, and down regulated genes that are associated with apoptosis, iron handling and immune response. There is, however, a dichotomy as, whilst HIV-1 group O and HIV-2 ETP share in common the majority of core elements in the top 20 ranking genes, the highest ranking core response genes regulated in response to HIV-1 subtype B are generally not shared. It could be that this indicates that the core response,

Table 4.3 Table showing core response genes that are in the top 20 up and down regulated genes, in ST1-R5 cells infected with HIV-1 subtype B, group O and HIV-2 ETP.

UP REGULATED – CORE RESPONSE				DOWN REGULATED – CORE RESPONSE			
Gene	B	O	ETP	Gene	B	O	ETP
Vascular endothelial GF	N*	Y*	Y	MHC class I	N	Y	Y
HMG CoA reductase	Y	Y	Y	B cell RAG assoc protein	N	Y	Y
7-DHCR	N	Y	Y	Ferritin	Y	Y	Y
acetyl CoA thiolase	N	Y	Y	annexin A2	N	Y	Y
mevalonate kinase	N	Y	Y	Pirin	N	Y	Y
AP2	N	Y	Y	CD27 Siva BP	Y	Y	Y
AP2 assoc. protein	N	Y	Y	KIAA0368 protein	N	Y	Y
cyclin G2	Y	N	N	HIV Tat interactive protein	N	N	Y
splicing factor arg/ser rich	Y	N	N	FXYD domain containing ion transport regulator 2	N	N	Y
				myosin light pp	N	N	Y
				Hypothetical protein FLJ10432	N	N	Y
				Gluathione-s-transferase pi	Y	N	N

*Y/Y indicates that the core response gene in question is in the top 20 up/down regulated genes
*N indicates that the core response gene in question is NOT in the top 20 up/down regulated genes

in subtype B infection, is modulated in a different way compared to HIV-2 and group O infections. This observation, however, may also be a product of the way in which data was analysed. As described in section 4.3.2.1-2, greater variation in the data for subtype B arrays at 0 HPI affected the way in which values of 'significance' were calculated in SAM. It may therefore be that the top 20 ranking genes for subtype B do not represent only the most strongly up and down regulated genes 24 HPI, as they do for group O and HIV-2 ETP, due to the described experimental 'noise'.

In order to determine, therefore, whether the top 20 significantly regulated genes in response to subtype B infection are dissimilar to those regulated by O and ETP by virtue of gene expression or if it is an artefact of data processing, the log (2) gene expression ratios for each of the 9 up- and 12 down regulated genes (Table 4.3) at 0 and 24 HPI were analysed (Figure 4.16). In 8 of the 9 up regulated genes, levels of gene expression account for the results seen. For example, HMG Co-reducatase is highly up regulated in response to all viruses and this gene is found in the top 20 for each. Similarly, for 9 of the 12 down regulated genes the levels of gene expression account for the results seen; for example for CD27 binding protein (SIVA) and ferritin there is a clear and consistent down regulation of gene expression, which is reflected in the ranking analysis. In the case of VEGF, 7-dehydrocholesterol reductase, acetyl-Coenzyme A acetyltransferase, mevalonate kinase, adaptor-related protein complex 2 and transcription factor AP2 gamma, where these core response genes are in the top 20 for

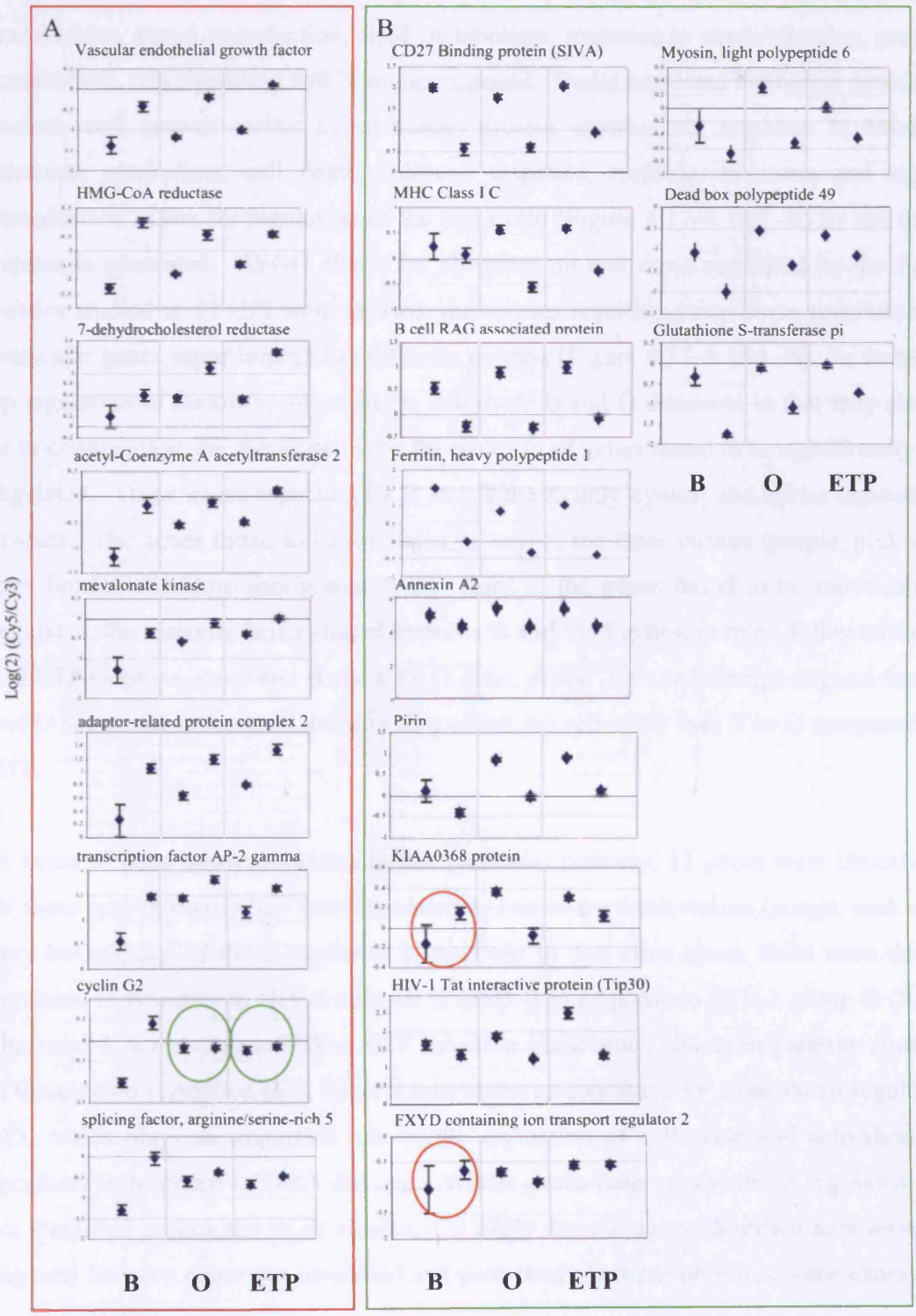
O and ETP but not B, however, the graphs indicate that the changes in gene expression in response to subtype B are comparable to those seen in response to O and ETP, if not stronger. In these cases the higher level of variation in the subtype B data at 0 HPI has lead to the underestimation of the significance of variation, in comparison to the other virus conditions. With regards to the cyclin G2 data, this gene is found to be highly significantly regulated by SAM in response to subtype B, but less so in response to O or ETP infection. However, by looking at the raw data it might be presumed that this gene is not significantly regulated in response to O and ETP. In this case, therefore, the opposite to the situation with subtype B could be true in that there is so little variation at 0 HPI in O and ETP data that the small variation between 0 and 24 HPI is found to be significant. Alternatively, these results may be false, within the 5% FDR of the SAM method.

For the down regulated genes, a similar picture emerges. The data for FXYD domain containing ion transport regulator 2 (FXYD2) and KIAA0368 protein, however, are unusual in that in response to B infection these genes appear to be up regulated, but to O and ETP they are down regulated. Calculation of p-values from the log(2) data for KIAA0368 expression at 24 HPI compared to 0 HPI revealed that the apparent up regulation in response to subtype B is not significant ($p = 0.079$), whereas the down regulation in response to O and ETP is ($p = 0.00123$ and 0.0169 respectively). The error bars for B at 0 HPI indicates that this data should be treated with caution due to excessive variability. In the case of FXYD2, calculation of p-values reveals that there is no significant difference in expression of this gene as a result of B, O or ETP infection ($p = 0.199$, 0.0617 and 0.458 respectively). Again, this gene may be one of the 5% incorrectly called as being significantly regulated, within the 5% FDR. This analysis provides a useful quality control for SAM: all genes of interest should be analysed in this way to ensure exclusion of false positives/inclusion of false negatives.

4.3.2.7 Genes regulated by 2/3 viruses

The core response of ST1-R5 cells to HIV infection has been defined as the genes significantly up and down regulated, in response to all viruses studied, at 24 HPI. As shown in Figure 4.12, however, there are many genes significantly regulated at 24 HPI that were found in response to only two viruses. To investigate these responses further

Figure 4.16 Analysis of core response genes that are in the top 20 up (A) and down (B) regulated genes, in ST1-R5 cells infected with HIV-1 subtype B, group O and HIV-2 ETP. The data plotted is the mean Log(2) (Cy5/Cy3), and standard error, of 6 replicate spots for each gene 0 and 24 HPI.



lists of genes positively and negatively regulated in two out of the three infections studied were compiled and placed into biological process categories, as defined by the Gene Ontology Consortium (http://apps1.niaid.nih.gov/david/). This showed that these genes essentially fall into the same categories identified previously for the core response, including up regulation of genes associated with cell cycle arrest, nucleic acid metabolism, signal transduction, lipid metabolism, response to stress/stimulus, protein metabolism, cell signalling and immune response. Down regulated biological processes include cell growth and/or maintenance, protein metabolism, response to external stimulus, catabolism, cell death, immune response, response to stress and signal transduction. Data for regulation of the cell cycle (Figure 4.17-A and -B) by the three viruses is presented. KEGG charts for all genes up and down regulated by the three viruses studied at 24 HPI were created, the viruses resulting in up/down regulation of particular genes super-imposed as different colours (Figure 4.17-A and -B). In terms of up regulation of factors involved in the cell cycle B and O dominate in that they alone, or in combination, are associated with the majority of genes found to be significantly up regulated. These genes regulated by B and O are mainly cyclins and cyclin-dependent kinases. The genes found to be regulated by two of the three viruses (purple, pink and grey boxes) constitute approximately one third of the genes found to be significantly regulated, the majority being shared between B and O (5 genes, purple), followed by O and ETP (3 genes, grey) and B and ETP (1 gene, pink). This order might suggest that B and O are closer in terms of the way they affect the cell cycle than B or O compared to ETP.

In terms of gene down regulation in this particular pathway, 11 genes were identified. Of these approximately half were regulated by two of the three viruses (purple, pink and grey boxes), and of those regulated in response to one virus alone, three were down regulated in response to HIV-1 subtype B (red), 3 in response to HIV-1 group O (light blue) and 1 in response to HIV-2 ETP infection (dark blue). No clear patterns emerge in these down regulation data, but it is interesting to note that ETP alone down regulates p53, which plays an important role in the regulation of cell cycle and activation of apoptosis in response to DNA damage. Whilst genes both up and down regulated are not identified in response to all viruses, it is likely that the genes identified here serve to augment the core responses identified and described. It is possible that some examples

Figure 4.17 - A. KEGG chart of genes associated with the cell cycle, significantly up regulated in ST1-R5 cells at 24 HPI with HIV-1 subtype B, group O and HIV-2 ETP.

The colours represent which infection leads to up regulation of the gene in question ie. TGFß is up regulated during infection with B and ETP (pink) and Kip1,2 is up regulated during infection with subtype B alone (red).

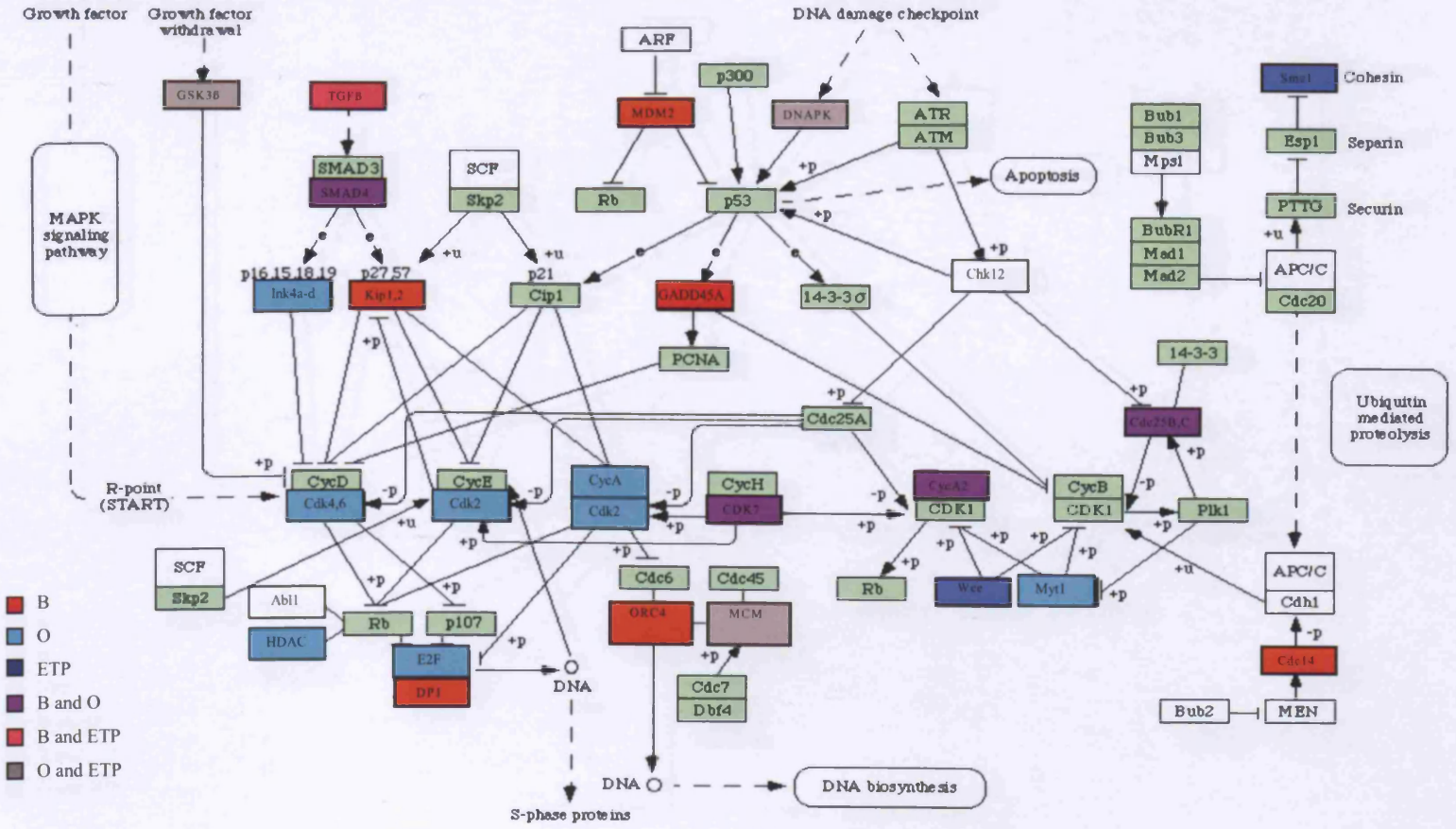
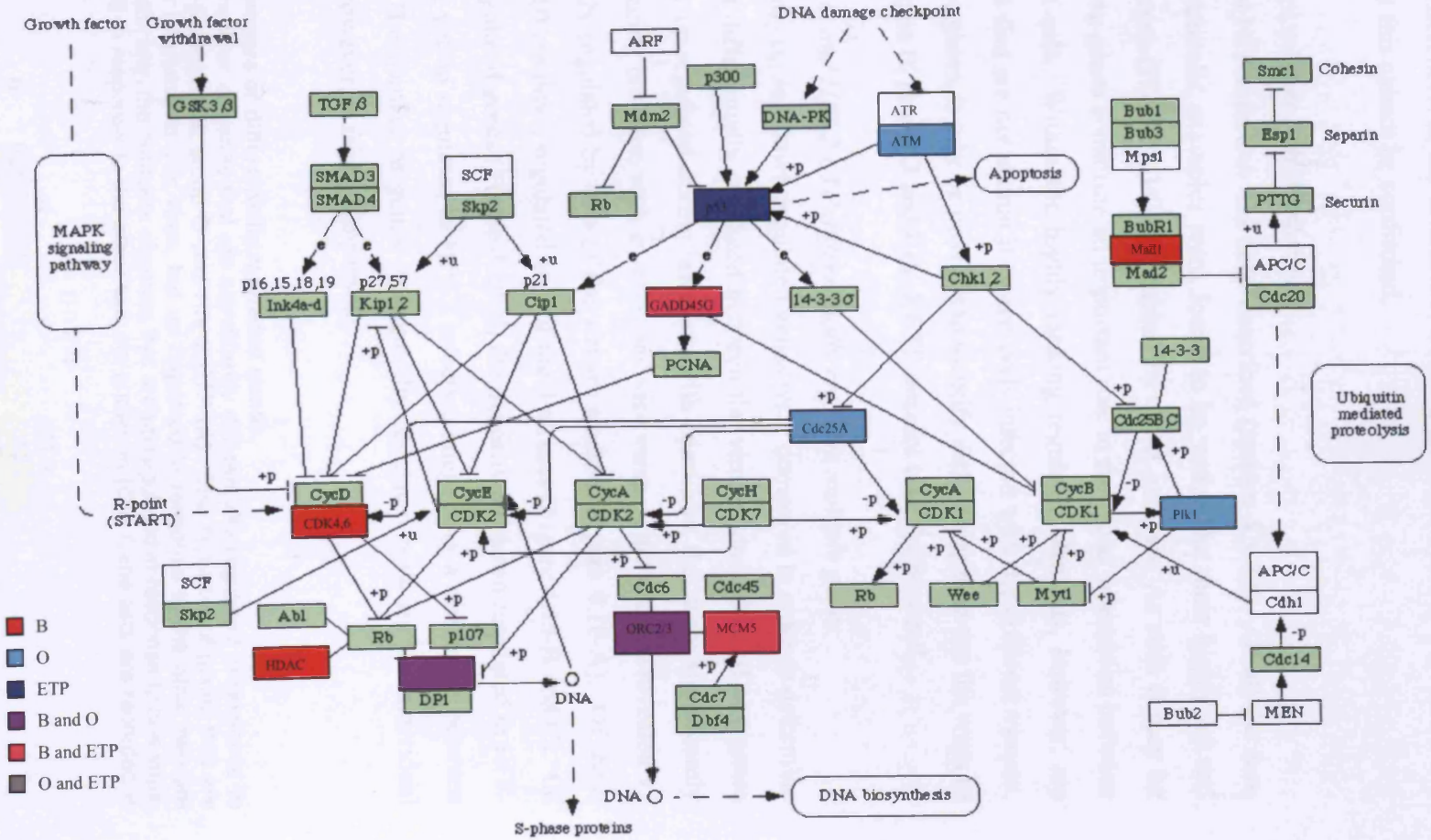


Figure 4.17 - B. KEGG chart of genes associated with the cell cycle, significantly down regulated in ST1-R5 cells at 24 HPI with HIV-1 subtype B, group O and HIV-2 ETP.

The colours represent which infection leads to down regulation of the gene in question ie. GADD45G is down regulated during infection with B and ETP (pink) and HDAC is down regulated during infection with subtype B alone (red).



represent subtle differences in the interaction between the host cell and different HIV types/subtypes, but this cannot be confirmed.

4.3.2.8 Diverging responses to HIV infection

A core response to HIV infection has been described (section 4.3.2.5). Of all the core response genes, particular examples were found to be within the most highly up and down regulated genes (Figure 4.16), as defined by SAM analysis. As such it may be concluded that these genes constitute an important role in the 'core' interaction between HIV and the host cell. Within the highly ranking regulated elements, however, are examples of genes that are not shared between cells infected with the different viruses. In examining these genes, it may be possible to identify differences between the ways in which HIV-1 subtype B, group O and HIV-2 ETP interact with the host cell.

HIV-1 subtype B, O and HIV-2 ETP differentially regulate multiple genes

Lists of significantly up and down regulated genes were compared in order to determine whether any were differentially regulated between the viruses studied i.e. if any genes were significantly up regulated during infection with one virus, that were significantly down regulated during infection with another, and vice versa. This analysis revealed 95 genes differentially regulated by two of the viruses studied (Figure 4.18-A). Of these four are up- and 10 are down regulated in 2 of the 3 viruses (Figure 4.18-B and C). Of these 10 down regulated genes (Figure 4.18-C) the majority are down regulated in HIV-1 B/O infections, yet up regulated in HIV-2 infection suggesting a difference between HIV-1 and -2. The number of genes differentially regulated between the individual viruses studied, however, is fairly comparable.

Figure 4.18 **Numbers of differentially regulated genes.**
Table shows the number of genes that are significantly differentially regulated in response to infection with HIV-1 subtype B, group O and HIV-2 ETP **(A)**. The numbers of genes that are down regulated in response to one virus, but up regulated in response to the other two are shown in **(B)**. Conversely, the numbers of genes that are up regulated in response to one virus, but down regulated in response to the other two are shown in **(C)**. Gene lists are recorded in Appendix, Table 3.

DOWN

UP	B	O	ETP
B	A	19	12
O	17		5
ETP	27	15	

UP	B	O	ETP
BO	B		1
OETP	2		
ETPB		1	

UP	BO	OETP	ETPB
B	C	2	
O			X
ETP	8		

In order to further dissect the nature of these differences in gene expression during infection with HIV-1 subtype B, group O and ETP, of the 95 differentially regulated genes that have been identified, only those genes that fall within the top 50 significantly regulated (as defined by d-value) will be considered. There are 10 genes, including three that indicate a difference between HIV-1 and HIV-2 in terms of cell cycle regulation and sensitivity to apoptosis, five indicating differences in transcriptional regulation, and two in relation to cell signalling.

Cell cycle regulation and sensitivity to apoptosis

LAG1 was ranked at position 20 in the list of genes up regulated during HIV-2 ETP infection, but was ranked 602 in the list of genes down regulated during HIV-1 group O infection. Over expression of LAG1 has been shown to selectively induce the synthesis of stearoyl-containing sphingolipids and ceramide in mammalian cells, and LAG1 and its homologues are likely to play a role in ceramide signalling, which affects growth, proliferation, stress resistance, and apoptosis (Jazwinski and Conzelmann, 2002). It may be postulated, therefore, that an increase in LAG1 expression during HIV-2 infection potentiates an increase in ceramide synthesis, leading to one of cell cycle arrest, apoptosis, terminal cell differentiation or senescence, all representing forms of stress response. Interestingly, elevated intracellular levels of ceramide have also been shown to inhibit HIV-1 infection (Finnegan *et al.,* 2004), which was proposed to occur as a result of perturbation of localised membrane domain structure and organisation. Infectivity is reduced due to the low probability of CD4 and coreceptor engagement, and trafficking of virions via an endocytic pathway that leads to non-productive infection (Schaeffer *et al.,* 2004, Finnegan *et al.,* 2004). It would seem, therefore, if ceramide biosynthesis is differentially regulated during HIV-1 and HIV-2 infection, via LAG1, that this could contribute significantly to the differences in the way these viruses behave both *in vitro* and *in vivo.*

Secondly, Pim-2 oncogene, a pro-survival kinase which phosphorylates the pro-apoptotic protein BAD thus reversing Bad-induced cell death (Yan *et al.,* 2003) was ranked at positions 17 and 3 in the list of up regulated genes, in response to O and ETP infection, respectively, but was at position 174 the list of genes down regulated during subtype B infection. A reduced frequency of HIV-2 infection-associated *in vitro* apoptosis in T-cells has been reported (Machuca *et al.,* 2004). No explanations were

given for this finding, so perhaps the strong up regulation of Pim-2 oncogene during HIV-2 infection may bear further inspection. It is worth noting that in studies of mammalian reoviruses, which have served as useful models for studies of the viral and cellular mechanisms that are operative in host cell damage and death (Clarke and Tyler, 2003), Pim-2 is one of the few apoptosis regulatory genes that is down regulated following reovirus infection (DeBiasi *et al.,* 2003), associated with an enhancement in apoptosis. This provides support for the notion that differential regulation of PIM-2 between different genetic variants of HIV might engender these different viral variants with distinct apoptotic phenotypes. This consequently may be reflected in the cytopathology of the diseases they cause, a suggestion also put by the authors of a recent publication describing a decreased level of T-cell apoptosis during HIV-2 infection *in vitro* (Machuca *et al.,* 2004).

Finally, two individual ESTs corresponding to cyclin-dependent kinase 5 (cdk5) were found to be down regulated in response to HIV-1 subtype B infection (ranked at 16 and 17) and HIV-1 group O infection (ranking 48 and 53), but were up regulated in response to HIV-2 ETP (ranking 433) infection. Cdk5 has been shown to influence several cellular processes in terminally differentiated cells, in particular neurons and in the kidney, and HIV-1 Tat has been shown to down-regulate the expression of p35, a neuron-specific activator of cdk5. This leads to deregulation of neuronal differentiation and survival (Peruzzi *et al.*, 2002), suggesting a role for factors released from HIV-1 infected cells that trigger a cascade of events leading to neurodegeneration. Furthermore, a role for Cdk5 as a regulator of podocyte differentiation, proliferation, and morphology has been demonstrated *in vitro* and *in vivo*: specific inhibition of Cdk5 in differentiated cultured podocytes induces shape changes, with cellular elongation and loss of process formation compared to the characteristic phenotype (Griffin *et al.,* 2004). HIV-associated nephropathy (HIVAN), characterised by proteinuria and progressive renal failure, is a well-known complication of HIV infection, and reversal of HIVAN end-stage renal failure after the initiation of HAART has been demonstrated (Scheurer, 2004). Thus, HIV-1 associated down regulation of cdk5, via Tat, may contribute to the development of HIVAN in HIV-1 infected patients. Its differential regulation during HIV-2 infection indicates that differences between these viral variants may influence the frequency of certain HIV-associated pathologies. The significance of differential cdk5 regulation via Tat within T-cells, however, is not clear.

Differential activation of cdk5, together with Pim-2 and LAG1, may therefore indicate differential modulation of cellular survival signals in HIV-1 versus HIV-2 infection, potentially impacting on the pathology of these diseases.

Transcriptional regulation

TFIIH, which interacts with HIV-1 Tat as a component of the HIV-1 transcription pre-initiation complex (Garcia-Martinez *et al.,* 1997) was ranked at position 46 in the list of down regulated genes, in response to ETP infection, but was at position 101 in the list of genes up regulated during subtype B infection. In addition, cyclin dependent kinase 7 (Cdk7), a component of CAK (a trimeric complex of cyclin H, MAT1 and Cdk7 which is an essential component of the TFIIH transcription factor), was significantly up regulated in response to subtype B and group O infection (Nekhai *et al.,* 2000). Together, these data suggest a difference between HIV-1 and HIV-2 in the expression of host factors essential for viral transcription: in HIV-1 infection there is a greater up regulation of host factors required for viral transcription, compared to HIV-2 infection. Similarly, general control of amino-acid synthesis 5-like 2 (hGCN5) was ranked at position 39 in the list of up regulated genes in response to ETP infection, but was at position 122 the list of genes down regulated during subtype B infection. Research has shown that Tat is directly acetylated at lysine 28 within the activation domain, and lysine 50 in the TAR RNA binding domain, by Tat-associated histone acetyltransferases p300, p300/CBP-associating factor, and hGCN5 (Col *et al.,* 2001, Bres *et al.,* 2002) so differential regulation of hGCN5 between HIV-1 subtype B and HIV-2 ETP might suggest a different relationship between HIV-1 and –2 Tat proteins, and the host factors that activate them. In addition, hGCN5 has been shown to be involved in histone acetylation and factor recruitment at the HIV LTR in response to Tat, so differential regulation of hGNC5 may also indicate divergent processes of LTR activation between HIV-1 and HIV-2 (Lusic *et al.,* 2003). Furthermore, B-cell CLL/lymphoma 6 (BCL6) is a sequence-specific repressor of transcription and was ranked at position 27 in the list of down regulated genes in response to ETP infection, but was at position 238 in the list of genes up regulated during subtype O infection. BCL6 can repress transcription from HIV-1 promoter/enhancer region (Baron *et al.,* 1997), possibly explaining its up regulation during HIV-1 O infection. BCL6 may also regulate apoptosis by repression of programmed cell death-2 (PDCD2) (Baron *et al.,* 2002).

Between HIV-1 subtype B and group O infections, retinoid X receptor alpha (RXRA), a nuclear receptor, was ranked at position 37 in the list of down regulated genes in response to group O virus, but was at position 653 in the list of genes up regulated in response to subtype B. It was recently demonstrated that retinoids, the bioactive metabolites of vitamin A, repress HIV-1 replication in monocytic cell lines by blocking LTR-directed transcription by transactivation of cellular gene expression (Hanley *et al.*, 2004). The finding that the synthetic pan-retinoic acid receptor antagonist BMS-204493 activates HIV-1 replication in a dose-dependent manner, implies that modulation of the retinoid receptor during infection may have an important effect on regulation of infection.

Cell signalling

Mitogen-activated protein kinase 8 (MAPK8/JNK) is involved in a wide variety of cellular processes such as proliferation, differentiation, transcription regulation and development, and is known to be activated by HIV-1 Tat (Kumar *et al.*, 1998) and Nef (Varin *et al.*, 2003); this activation having been postulated to contribute to AIDS pathogenesis by fuelling the progression of disease via stimulation of HIV-1 provirus present in such cellular reservoirs as mononuclear phagocytes. Analysis of differentially regulated genes revealed that MAPK8 was ranked at position 29 in the list of up regulated genes in response to ETP infection, but was at position 538 in the list of genes down regulated during subtype B infection, suggesting differences between HIV-1 and HIV-2 in the manner in which such intracellular signalling networks are modulated during infection.

Finally, chloride channel 4 (ClC4) was ranked at position 34 in the list of up regulated genes in response to ETP infection, but was at position 468 in the list of genes down regulated during group O infection. This channel is involved in the regulation of hepatic copper transport (Wang and Weinman, 2004), and whilst other physiological roles remain unclear, defects it this channel have been associated with myotonia and certain kidney disorders (Dworakowska and Dolowy, 2000). Recent molecular investigation, however, has indicated that ClC4 contributes to endosomal acidification and trafficking: specific disruption of endogenous ClC4 expression by transfection of ClC4 antisense cDNA acidified endosomal pH and altered transferrin trafficking in cultured epithelial cells (Mohammad-Panah *et al.*, 2003). The modulation of this ion

channel, during infection, may therefore be of importance as HIV-1 Tat has been shown to enter T-cells using clathrin-mediated endocytosis, before low-pH-induced and Hsp90-assisted endosomal translocation, following which cell responses are induced from the cytosol (Vendeville *et al.,* 2004). Down regulation of this channel during HIV-1 group O infection may therefore reflect a mechanism of altering Tat delivery and the resulting bystander effect (Vendeville *et al.,* 2004). Furthermore, the entry, trafficking and exit of HIV particles from cells are heavily dependent on the endosomal compartment. As mentioned previously, in addition to assembly at the plasma membrane, retroviruses utilise late endosomal membranes/multivesicular bodies as assembly sites, implying an endosome-based pathway for viral egress (Sherer *et al.,* 2003). Thus, differential regulation of ClC4 in response to HIV-1 and HIV-2 infection may indicate divergent ways in which these viruses interact with or modulate these pathways, to ensure their passage through the cell. These differentially regulated genes are summarised in Figure 4.19 and the raw Log(2) data for these genes are shown in Figure 4.20.

Figure 4.19. Summary of genes found to be differentially regulated in response to HIV-1 subtype B, group O and ETP infection.
The gene names are shown within the bars and the virus in which the opposite regulation was observed is noted next to the bar i.e. in response to subtype B infection one gene, cdk5, was down regulated for which the opposite regulation was observed in response to ETP infection (up regulated). The genes shown (in the bars) are only those within the top 50 ranking up or down regulated genes, detected 24 HPI.

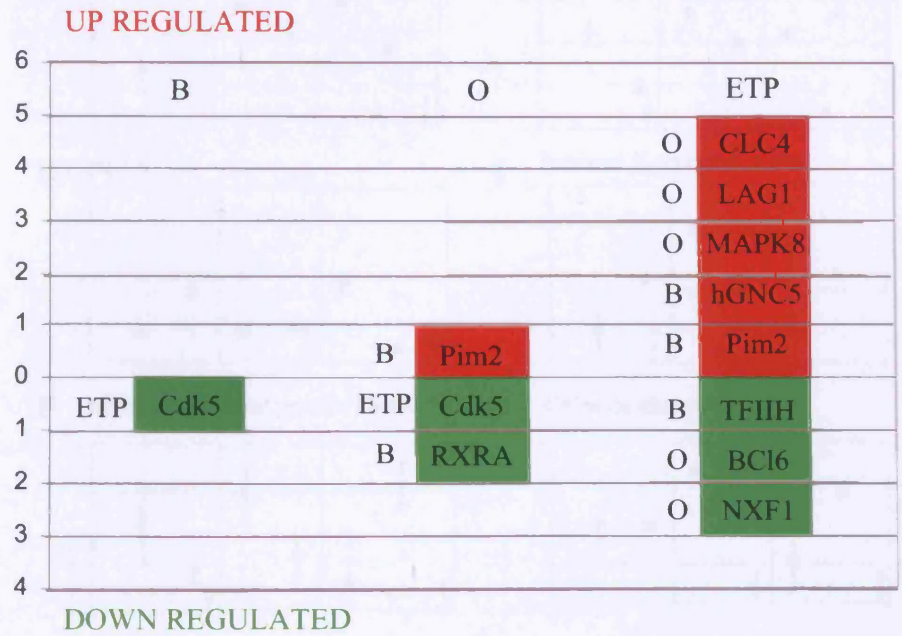
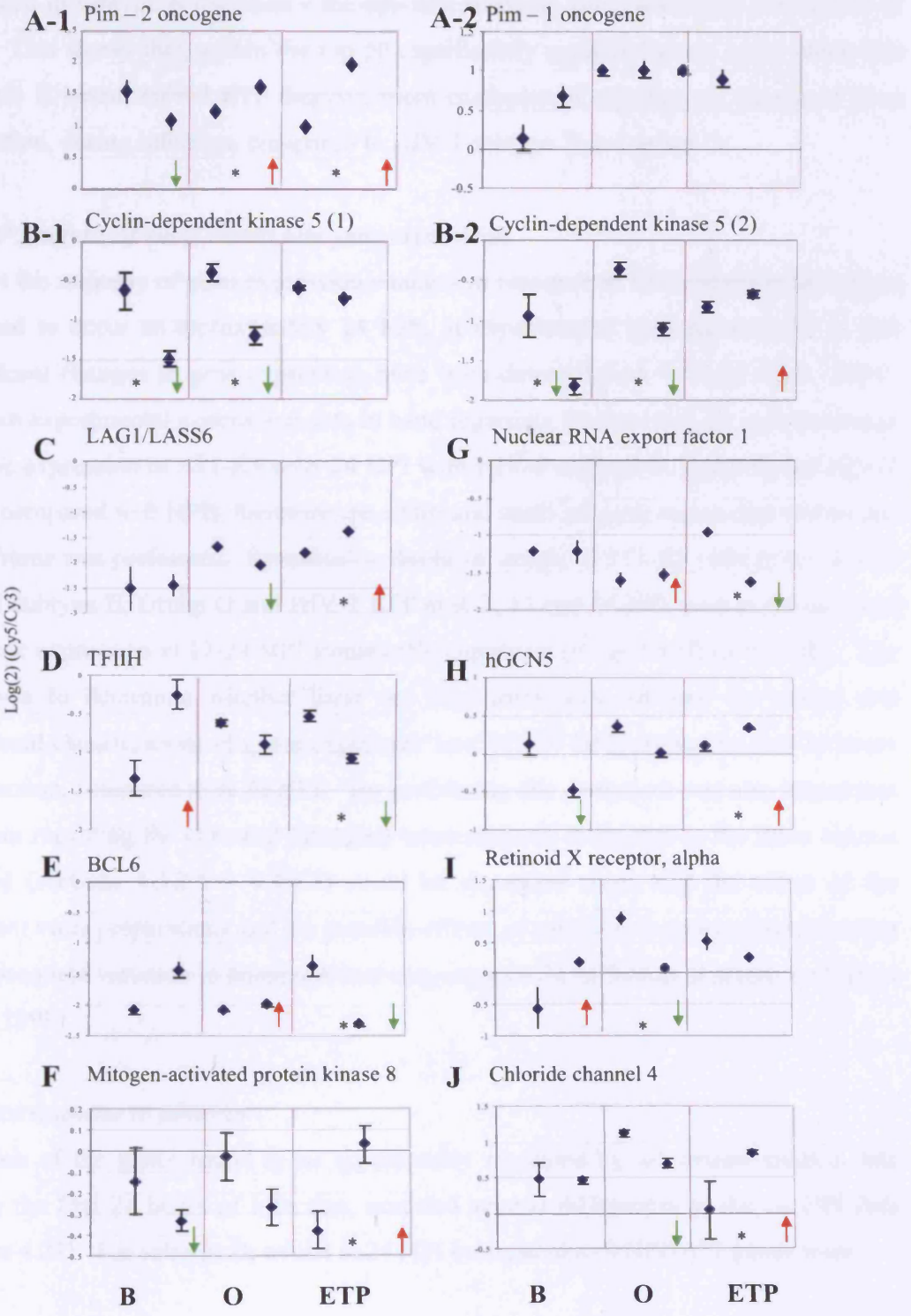


Figure 4.20 Analysis of genes for which significant diverging responses were identified, one of which is within the top 50 of either up or down regulated genes, in ST1-R5 cells infected with HIV-1 subtype B, group O and HIV-2 ETP.
The data plotted is the mean Log(2) (Cy5/Cy3), and standard error of 6 replicate spots for each gene 0 and 24 HPI. The virus in response to which the gene regulated was within the top 50 is marked in each case with an asterisk (*).



From the summarised data it is clear that ETP infection results in strong and significant up and down regulation of eight genes, that are significantly regulated in exactly the opposite way in response to HIV-1 subtype B (n=3) or group O (n=5) infection. Group O infection result in differential regulation of three genes, two of which the opposite was observed for subtype B and one in ETP infection, and one gene was differentially regulated in subtype B infection – the opposite response being identified in response to ETP. This shows that, within the top 50 significantly regulated genes upon which this analysis is based, HIV-2 ETP displays more examples of significantly divergent gene regulation, during infection, compared to HIV-1 subtype B and group O.

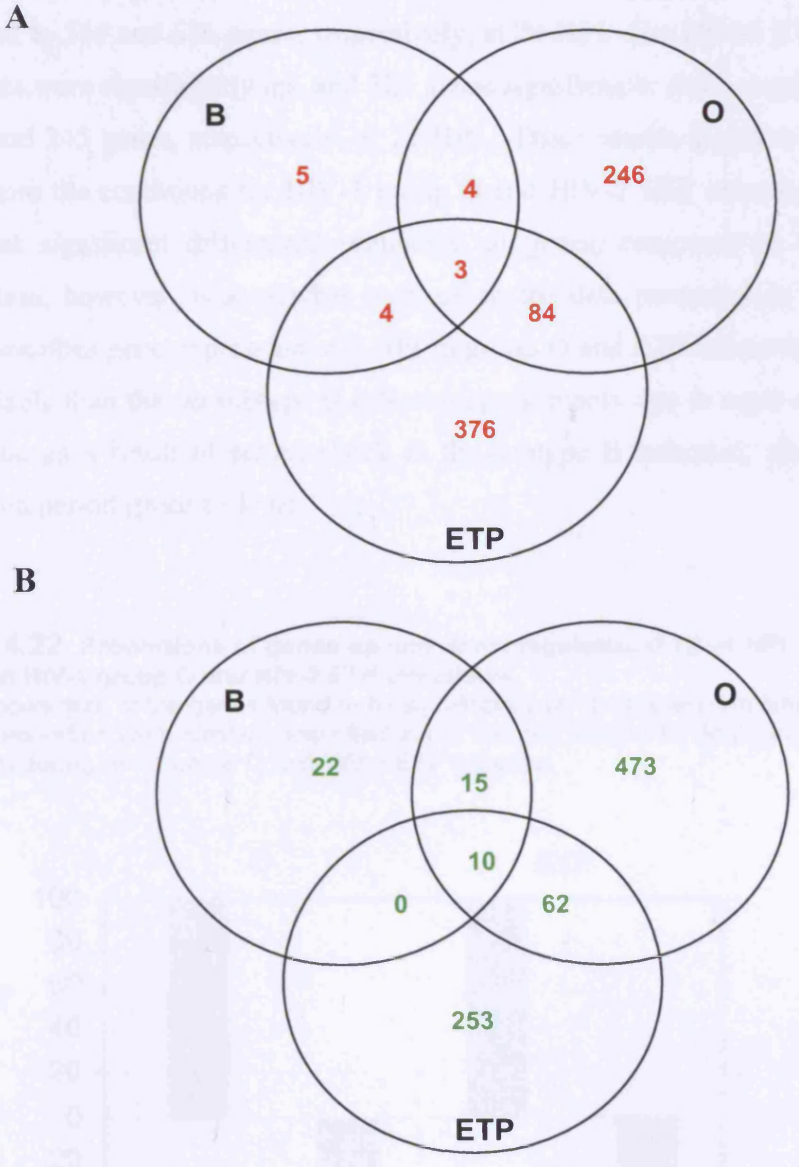
4.3.2.9 Analysis of early versus late gene expression

Whilst the majority of gene expression changes in response to HIV infection have been reported to occur at approximately 24 HPI, at experimental time points prior to this significant changes in gene expression have been detected (van 't Wout *et al.,* 2003). With an experimental system and data in hand regarding the core and divergent changes in gene expression in ST1-R5 cells 24 HPI with HIV-1 subtype B, group O and HIV-2 ETP (compared to 0 HPI), therefore, an additional study of gene expression within this time-frame was performed. Specifically, duplicate arrays of ST1-R5 cells infected with HIV-1 subtype B, Group O and HIV-2 ETP at 0, 2, 12 and 24 HPI were performed and the gene expression at 12-24 HPI statistically compared (using SAM) to 0-2 HPI. The aim was to determine whether there are any differences between the extent and functional classifications of genes expressed 'late' (12-24 HPI) during the first 24 hours of infection, compared to at 24 HPI. By performing this analysis it was also hoped that the data regarding the core and diverging transcriptional responses to the three viruses studied (sections 4.3.2.5 – 4.3.2.7) could be expanded upon, and the effect of the different virus preparations and the possible effects of serum re-exposure after infection on subsequent variation in transcriptional responses could be further characterised (Iyer *et al.,* 1999).

The late response to infection

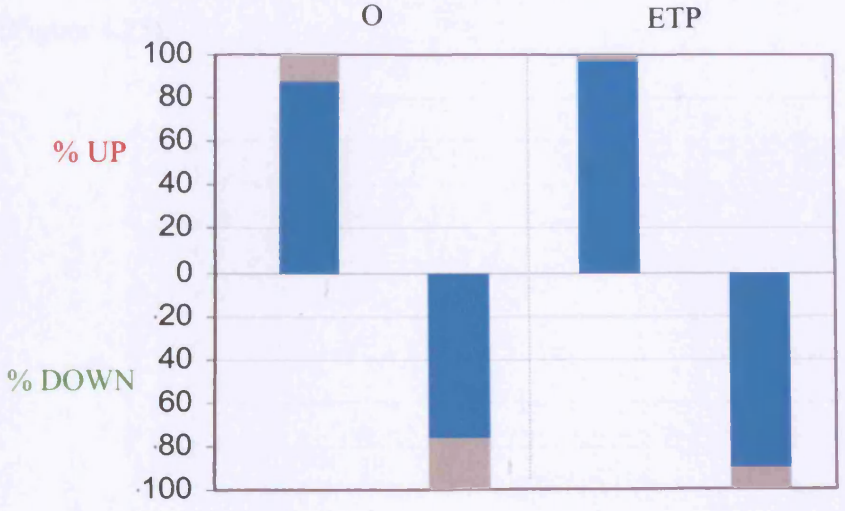
Analysis of the genes found to be significantly regulated by all viruses studied, late during the first 24 hours of infection, revealed several differences to the 24 HPI data (Figure 4.21). For subtype B, whilst at 24 HPI (compared to 0 HPI) 677 genes were

Figure 4.21 Core and diverging responses to HIV-1 subtype B, group O and HIV-2 infection at 12-24 HPI.
(A) Venn diagram showing numbers of significantly up regulated genes at 12-24 HPI, as compared to 0-2 HPI, in ST1-R5 cells infected with HIV-1 subtype B (B), group O (O) and HIV-2 (ETP). **(B)** Venn diagram showing numbers of significantly down regulated genes at 12-24 HPI, as compared to 0-2 HPI, in ST1-R5 cells infected with HIV-1 subtype B (B), group O (O) and HIV-2 (ETP).



found to be up regulated and 646 genes were down regulated, analysis of the 12-24 HPI data compared to 0-2 HPI, revealed only 16 up- and 47 significantly down regulated genes. This observation is consistent with previous observations that the majority of gene regulation in response to HIV infections occurs around 24 HPI (van 't Wout *et al.*, 2003). Analysis of gene expression changes in response to HIV-1 group O and HIV-2 ETP at 12-24 HPI, however, generated very different results. For group O at 12-24 HPI, 337 genes were significantly up- and 560 genes significantly down regulated, compared to 554 and 616 genes, respectively, at 24 HPI. For HIV-2 ETP at 12-24 HPI 467 genes were significantly up- and 325 genes significantly down regulated, compared to 317 and 245 genes, respectively, at 24 HPI. These results therefore indicate that in response to the conditions for HIV-1 group O and HIV-2 ETP infection, at 12-24 HPI there was significant differential regulation of genes, compared to 0-2 HPI. This observation, however, is somewhat contrary to the data presented in section 4.3.2.2, which describes gene expression at 0 HPI in group O and ETP infection as being much less variable than that in subtype B infection, presumably due to rapid changes in gene expression as a result of serum shock in the subtype B infection, during the 3 hour incubation period (prior to T=0).

Figure 4.22 Proportions of genes up and down regulated at 12-24 HPI, as compared to 24 HPI, in HIV-1 group O and HIV-2 ETP infections.
Graph shows that, of the genes found to be significantly up- and down-regulated at 24 HPI, only a small proportion were similarly identified in the merged 12/24 HPI data (grey sections of the blue bars) during HIV-1 group O and HIV-2 ETP infection.



In order to determine, therefore, if these changes in gene expression in HIV-1 group O and HIV-2 ETP infection 0-2 HPI were as a result of infection, an analysis of the proportions of genes commonly up- and down regulated between 12-24 HPI, compared to 24 HPI was performed. This showed that for group O, only 15-20% of the genes up- and down regulated at 12-24 HPI were also identified as being significantly regulated at 24 HPI, and for HIV-2 ETP a smaller number of genes (2-10%) were found in the analysis of both experiments to be similarly regulated (Figure 4.22). That is, the majority of genes identified as being significantly regulated in response to O and ETP between 12-24 HPI are not the same as those found to be significantly regulated at 24 HPI. This implies that these genes are not regulated in response to infection, but rather the conditions under which infection was performed. It is likely that the process of putting cells into complete medium at T=0, after infection in almost neat infected cell supernatant during the 3 h incubation period with HIV-1 group O and HIV-2 ETP, also initiated a form of serum shock response in these infections. In support of this hypothesis, a comparison of the genes up and down regulated at 0 HPI in response to subtype B infection to those up and down regulated at 0-2 HPI in group O and ETP infection revealed approximately 50% similarity. A model may therefore be proposed to explain these observations in which serum shock in the O/ETP infections occurs after infection, at the experimental T=0 (three hours after the serum shock in the subtype B infection). This delayed shock response in O/ETP infections gives the impression, when 0-2 HPI and 12-24 HPI data are compared, that significant changes in gene expression are taking place sooner after infection in response to these viruses, compared to subtype B. This result is, however, likely to be an artefact of the infection strategy (Figure 4.23).

Figure 4.23 Model of changes in gene expression in response to infection and infection conditions, over time.
This schematic indicates that cells infected with subtype B are subject to a brief serum shock at the beginning of the experiment (blue dashed line), whereas group O and ETP infections are exposed to serum in a sustained manner after 3 h (black dashed line). This produces the high variation in subtype B T=O data. Virally induced changes in gene expression do not appear until close to T=24 (complete blue and black lines for B and O/ETP respectively). When T=12-24 (B) and T=0-2 HPI (A) data are compared, therefore, the group O and ETP infections shown a greater number of genes differentially regulated at 12-24 HPI, due to the delayed serum exposure (at T=0). The differences in response to serum re-exposure depicted may be as a result of cells in the subtype B infection going from old medium to fresh medium, during infection, whereas the group O/ETP infected cells go from almost neat virus supernatant to complete medium (T=0), perhaps resulting in a more profound and lasting change in gene expression.

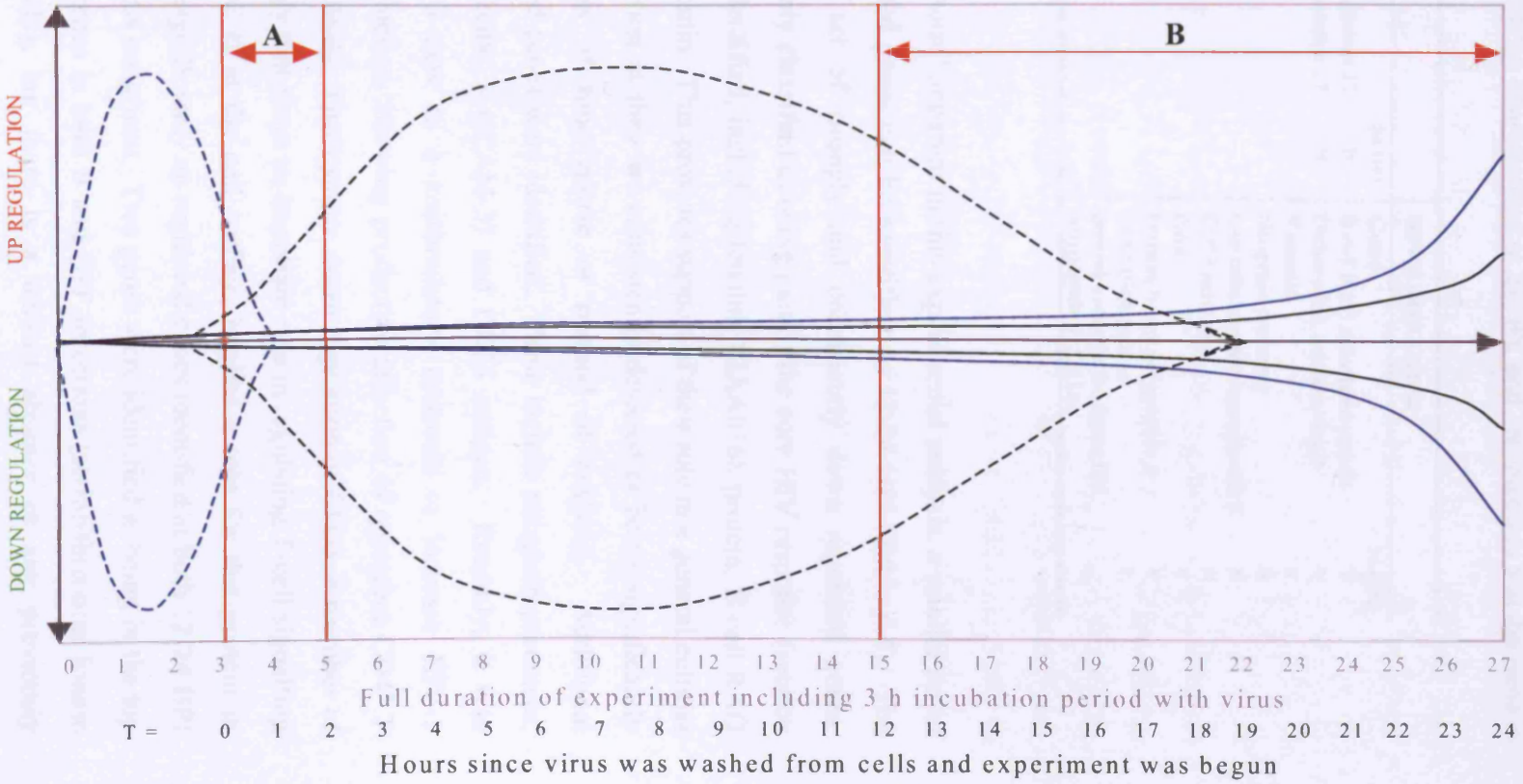


Table 4.4 Specific core response genes up and down regulated at 12-24 HPI.
These genes were identified in SAM analysis of 12-24 HPI vs. 0-2 HPI arrays, as being significantly regulated in response to all 3 viruses. 'Y' indicates that the gene in question was also identified in the core response characterised at 24 HPI, and '**N**' indicates that the gene in question was not.

UP REGULATED		DOWN REGULATED	
Gene	**24 HPI**	**Gene**	**24 HPI**
RNA binding motif protein 17	N	B cell RAG associated protein	Y
G protein-coupled receptor 17	N	Dishevelled, dsh homolog 2	Y
		Vimentin	Y
		Neogenin precursor	N
		Intercellular adhesion molecule 3	N
		CD53 antigen	N
		Pirin	Y
		Ferritin, heavy polypeptide 2	Y
		KIAA0368 protein	Y
		general transcription factor IIH, polypeptide 4, 52kDa	Y

The late core response

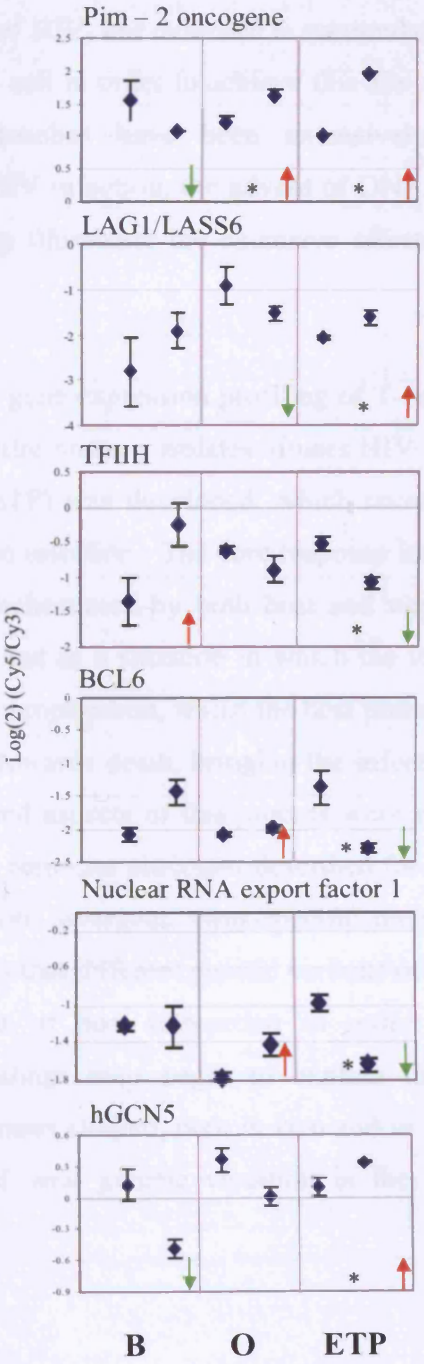
Despite the experimental 'noise' apparent in this experimental analysis, a smaller set of core up- and down regulated genes can be identified, at 12-24 HPI (Table 4.4). The genes identified reveal a set of strongly and consistently down regulated genes. Examples of genes previously classified as being part of the core HIV response (section 4.3.2.5) have again been identified, including ferritin, KIAA0368 protein, B cell RAG associated protein and vimentin. This provides support of their role in a general cellular response during HIV infection as they are consistently detected as being significantly down regulated, regardless of time course or method of analysis. Additional significantly down regulated genes were identified. These include neogenin precursor, intercellular adhesion molecule 3 (ICAM-3) and CD53 antigen. Recently, it was demonstrated that ICAM-3 acts as a costimulatory molecule to increase HIV-1 transcription and viral production, allowing productive infection of quiescent $CD4^+$ T lymphocytes (Barat *et al.*, 2004). Furthermore down regulation of CD53, a member of the tetraspanin protein family that plays an important role in regulating T-cell signalling when associated with MHC II at the cell surface, implies a role for this protein in response to infection. The significantly up regulated genes identified at both 12-24 HPI and 24 HPI, however, are less congruent. Two genes were identified as being in the top 50 significantly regulated genes in both B and ETP infections (serine/threonine kinase 18 and STAT1 respectively), but there is a distinct absence of any previously characterised core response genes. This indicates that the 12-24 HPI data for up

regulated genes is not as similar to the 24 HPI data, as it is in the data for down regulated genes. The general trend that down regulated genes are identified in 24 HPI and 12-24 HPI merged data, but up regulated genes are not, does suggest that down regulation of genes happens earlier in HIV infection than up regulation, although this has not previously been observed and may merely be a by product of infection conditions, rather than infection itself.

Diverging responses late in infection

In order to determine if genes identified to be differentially regulated in cells infected with HIV-1 subtype B, group O and HIV-2 ETP at 24 HPI, were also differentially regulated in the second set of time courses and arrays performed (for 0-2 vs.12-24 HPI analysis), the 24 HPI data from the repeat experiments was analysed. The same patterns of up and down regulation were found in the repeat arrays for six of the ten genes of interest, namely: Pim2 oncogene; LAG1; TFIIH; BCl6; NXF1 and hGCN5, verifying this data (Figure 4.24). Results for the other four genes were not conclusive. These data therefore corroborate the findings from previous analysis of differentially expressed genes at 24 HPI (section 4.3.2.8, Figure 4.19): regulation of these genes are strongly associated with subtype/type specific infection.

Figure 4.24 Identification of genes differentially regulated late during the first 24 hours of infection between HIV-1 Subtype B, Group O and HIV-2 ETP.
Genes for which significant diverging responses were previously identified, one of which is within the top 50 of either up or down regulated genes, in ST1-R5 cells infected with HIV-1 subtype B, group O and HIV-2 ETP. The data plotted is the mean Log(2) (Cy5/Cy3), and standard error of 3 replicate spots for each gene 0 and 24 HPI. The virus in response to which the gene regulated was within the top 50 is marked in each case with an asterisk (*).



4.4 Discussion

HIV infects CD4+ T lymphocytes eventually inducing their severe depletion, which is the defining feature of AIDS. It is not clear precisely how the virus exploits the host cell to maximise viral particle production, or whether this process differs between different genetic variants of HIV, but evidence is accumulating that HIV plays a central role in regulating the host cell in order to achieve this aim (Arendt and Littman, 2001). While biochemical approaches have been extensively employed to study the intracellular response to HIV infection, the advent of DNA microarrays has provided a powerful new tool to help illuminate the extensive effects of HIV on host-cell gene expression.

In this study, a method of gene expression profiling of T-cells during the first 24 hours of infection with the low titre primary isolates viruses HIV-1 subtype B (SF162), group O (BF306) and HIV-2 (ETP) was developed, which reveals both core and diverging transcriptional responses to infection. The core response identified represents a general HIV-response program, orchestrated by both host and virus. The complex interplay between the two is borne out as a situation in which the virus attempts to harness the host cell to ensure its own propagation, whilst the host possesses mechanisms to disable the virus and drive itself towards death, bringing the infection cycle to an end. Many previously well-documented aspects of this process were observed in this study, with the addition of novel core response processes described for the first time. Augmenting this core response are more divergent, virus-specific transcriptional responses. The detection of these indicates that different genetic variants of HIV have acquired specific and directed mechanisms of host interaction in order to successfully propagate themselves. Several findings may begin to explain the observed differences in phenotype between the viruses studied, both *in vivo* and *in vitro*. The results presented outline the importance of viral genetic variation in the processes of infection and disease.

4.4.1 A method for microarray analysis of low titre, high multiplicity HIV infection.

For microarray analysis of HIV-1 subtype B, group O and HIV-2 ETP a high multiplicity of infection was required to ensure synchronous infection. Initial investigations of the interaction between HIV and T-cells were particularly hampered by the use of a low multiplicity of infection and long time courses (Geiss *et al.*, 2000). As the infectious titre of the primary isolate stocks used was low compared to that which can be achieved with cell line adapted viruses or molecular clones, to achieve the desired high multiplicity of infection a small number of cells were infected, meaning that RNA available from infected cells became limiting. Accordingly, a number of RNA extraction and amplification protocols were studied in order to identify a method which would effectively amplify RNA from 10^4 cells whilst preserving the relative abundances of different mRNAs within the starting population in a reproducible manner.

The resulting method proved to be efficient in producing linearly amplified RNA (with good correlation with un-amplified RNA) in a reproducible and consistent manner. Amplification of reference RNA was also performed in keeping with recent reports that it is important for sample and reference RNA to be generated in the same way (Li *et al.*, 2004b). The hybridisation properties of the amplified reference RNA were consistent with that of the un-amplified RNA, important as one of the main purposes of using a universal reference RNA is to enable different array experiments to be cross compared, provided they use the same reference RNA.

The study of low titre viruses such as those used here would previously have not been possible without the optimisation of the infection, RNA extraction and amplification protocols described. It should be noted, however, that there is a particular limitation inherent in the experiments here presented, relating to the comparison of viruses of different viral titre. Specifically, in order to equalise the volume in which infection was performed, different volumes of virus and medium were combined, for each isolate. As a result, during infection the cells were not only exposed to different viruses, but also different concentrations of cytokines, growth and signalling factors, which also have the potential to significantly affect cellular gene expression. This is clearly outlined in

sections 4.3.2.2 and 4.3.2.9 where extensive analysis was used to show that the infection conditions for HIV-1 subtype B compared to group O and HIV-2 ETP result in different patterns of gene expression early during infection, seemingly independent of the type of HIV infecting the cells. As such this study provides both an example of experimental variation potentially providing misleading data, and how to extract meaningful data from 'noisy' data using the appropriate analytical tools with due care and attention. The successful use of this method therefore demonstrates its utility in studying low titre viral infections of small numbers of cells, and may equally be applied to other viruses or low passage HIV isolates derived from patient samples. The use of purified virus may bear further investigation for future studies.

4.4.2 The core response to HIV infection

Genes identified as being significantly regulated during infection with all three viruses included those involved in modulation of cell growth and maintenance; transcription, transcriptional regulation and translation; response to external stimulus; cell signalling; immune responses; lipid and protein metabolism; and intracellular trafficking (section 4.3.2.5). It seems that this broad range of responses are elicited around 24 HPI and that whilst some may represent more universal cellular responses to infection others are more pointed. These may be specific methods of antiviral defence orchestrated by the host, for example down regulation of TIP30 which function as a Tat-enhancer, or are examples of gene regulation driven by the virus itself, such as down regulation of MHC class I-C by HIV Vpu (Kerkau *et al.*, 1997) and Nef (Andrieu *et al.*, 2001). It is very difficult when considering the majority of these core responses, however, to distinguish between the two forces - host and virus - when attempting to dissect the nuances of cellular transcriptional responses to HIV infection.

4.4.2.1 Virus versus host: driving the core response to HIV infection

The viral accessory proteins Tat, Nef, Vpr and Vpu have all been shown to affect the host in such a way to propagate infection, and the down regulation of MHC class I (Vpu/Nef), up regulation of negative cell cycle regulators (Vpr), and up regulation of growth factors (Tat) are all examples of their effects, detected in this study. Their functions, however, are most probably far from completely understood and the true extent to which they contribute to this core response is by no means clear. It would require the systematic knock out of these accessory genes in the context of infection to

ascertain their function in a given cell type. It is possible, however, that in the absence of one accessory gene another may compensate for it, or conversely loose certain facets of its functionality, further increasing the ambiguity regarding their function. Several studies have involved the exposure of cells to either endogenously or exogenously expressed accessory proteins such as Tat (Gibellini *et al.,* 2001, Izmailova *et al.,* 2003, reviewed in section 4.1.4.1), revealing important functional consequences of their presence in a host cell. Outside of the context of infection, however, such work can be difficult to interpret as the host is relatively inert: conclusions drawn are partly clouded by the lack of a host response which may significantly alter that which is elicited by the transient presence of a viral protein alone. Perhaps a better approach, therefore, may be to systematically knock out the genes being regulated by HIV accessory proteins, rather than the accessory proteins themselves, in the context of infection. There are several host genes identified as being regulated as part of this core response, which could be targeted in order to assess their significance during HIV infection.

4.4.2.2 VEGF

VEGF up regulation was found to be part of the core response to HIV infection, which has previously been attributed to the action of HIV Tat (Benelli *et al.*, 2000). The literature, however, is somewhat in disagreement as to the importance of this observation, in relation to HIV. Several studies have described an increase in VEGF expression levels during HIV infection both *in vitro* (Ascherl *et al.,* 1999) and *in vivo* (Wang *et al.,* 2004) - one study in particular associating elevation of serum VEGF with the occurrence of HIV encephalopathy (HIVE), the mechanism for which is not understood (Sporer *et al.,* 2004). Another study investigating the utility of VEGF concentrations in serum as a predictive marker for AIDS-related Kaposi's Sarcoma (AIDS-KS), however, found that serum concentrations of VEGF were not influenced by HIV-1 infection (Renwick *et al.,* 2002). The main importance of elevated serum VEGF has thus not been related to HIV infection itself, rather the focus has been on its role in the development of AIDS-KS. KS cells produce and respond to angiogenic factors such as VEGF, with this factor being crucial in the development of KS tumours (Cornali *et al.*, 1996). HIV Tat also plays a major role in the pathogenesis of AIDS-related KS by augmenting the angiogenic activities of VEGF, and activating the VEGF receptor-2 (Aoki and Tosato, 2003). Furthermore, KS cells themselves also display elevated expression of VEGF (Cornali *et al.*, 1996). As strong VEGF up regulation as a result of

HIV infection alone was observed in this study, however, it would be interesting to determine whether inhibition of VEGF expression will actually affect HIV replication. It has been proposed that VEGF up regulation via the activity of extracellular Tat is part of a viral-driven process of suboptimal activation of T cells, in which partial activation signals enhance latent HIV infection by favouring entry into anergy (Benelli *et al.,* 2000). Thus, RNA inhibition of VEGF expression or use of synthetic molecule inhibitors such as SU5416, which has been shown to decrease VEGF mRNA expression levels in a dose- and time-dependent manner (Zhong *et al.,* 2004), could be used to study infection in the absence of this factor. If VEGF expression is important to the process of HIV replication, as well as in the development of AIDS-related vasculopathies, this should become more clear.

4.4.2.3 FOS

Similarly, the Tat-mediated induction of fos described previously in uninfected T cells (Gibellini *et al.,* 2001) was detected as part of the core response to primary HIV infection. This has been proposed to be involved in the up regulation of genes critical for the activation of T lymphocytes, such as interleukin 2 (IL-2), thus contributing to an optimal environment for HIV-1 replication in activated CD4+ T cells. In addition to this function, however, the role of fos in the enhancement of HIV-1 gene expression in primary human CD4 T-cells, via NFAT-1, has been described (Chen *et al.,* 1998, Cron *et al.,* 2000, Giffin *et al.,* 2003). NFAT has been considered as a therapeutic target for prevention of HIV-1 LTR-directed gene expression in HIV infection (Cron *et al.,* 2000). Given the observation that fos is strongly up regulated as part of the core response to HIV infection, studies in which fos expression is inhibited and the efficiency of transcription and infection monitored, should help show how important this factor is: whether it is most important in priming the cellular environment for the process of viral replication, or whether its direct involvement in viral transcription is sufficiently influential to make it a valid therapeutic target. It has been observed that HIV-1 seropositive individuals are less likely to progress to clinical immunodeficiency while undergoing treatment with NFAT inhibitors such as cyclosporin A (Schwarz *et al.,* 1993, Levy *et al.,* 1995, Andrieu *et al.,* 1988), as part of combination therapy. Such inhibitors might help limit the risk of acquisition of resistance to other anti-viral agents. If fos were confirmed to play a significant role in the process of viral replication and be

important in the chronic immune activation characteristic of AIDS progression, it too could be a valid drug target.

4.4.2.4 TIP 30

TIP 30, a serine/threonine kinase which induces the expression of apoptosis related genes and interacts with amino acids 1-48 of HIV-1 Tat enhancing Tat activation of the HIV-1 LTR promoter (Xiao *et al.,* 1998), was down regulated as part of the core response to HIV. It would seem advantageous that an infected cell would down regulate TIP 30 in response to infection, as this may function as a means to both reduce Tat activity and cellular pathology as a result of an increase in apoptosis. Conversely, the reduction in expression of apoptotic factors could be advantageous to the virus. Experimentally, if TIP 30 were provided either exogenously in an experimental infection system, or effectively inhibited using a small molecule inhibitor or small interfering (si) RNA, it would be interesting to determine (as above for fos) if HIV transcription or infection efficiency would be altered. It is also interesting to note that expression of TIP 30 in an adenovirus vector system has been shown to inhibit the proliferation of tumour cell lines through both p53-dependent and p53-independent pathways (Zhang *et al.,* 2004), and that TIP 30 deficiency has conversely been shown to increase susceptibility to tumorigenesis (Jiang *et al.,* 2004). Thus, it may be that down regulation of TIP 30 during HIV infection may play a role in the complex processes that lead to the development of lymphoproliferative disorders (Little, 2003), as well as in the modulation of viral replication and the cellular microenvironment.

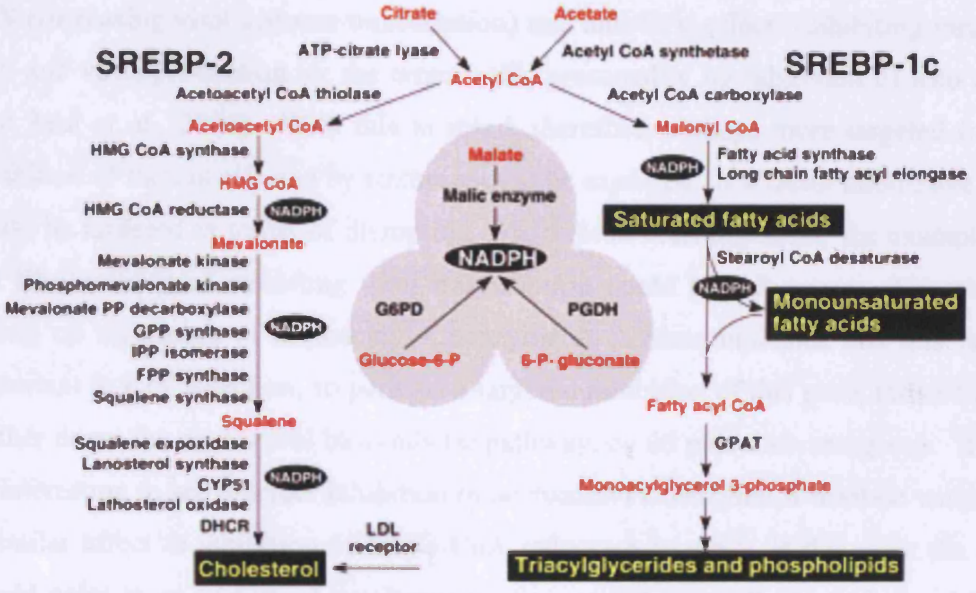
4.4.3 Identification of core response elements is clinically important

Whilst the core response to HIV infection may be taken as an indicator as to the most important and basic interactions between the host and virus, defining its replication strategy and pathological consequences, its characterisation may also be considered as an important tool in the search for future antivirals and biomarkers of disease. A good example of the latter was the identification of VEGF as being significantly up regulated in response to HIV-1 subtype B, group O and HIV-2. The biological significance of this was discussed in section 4.4.2.1 and the literature described: there is a lack of consensus regarding both the relationship between HIV and VEGF elevation, and its potential use as a biomarker in the prediction or therapeutic monitoring of AIDS-KS (Renwick *et al.,* 2002). Given the clear results in this study in terms of transcriptional

up regulation of VEGF, however, further experiments are required to confirm up regulation at the secreted protein level. If this could be shown, studies of patients would bear re-visiting. Previous *in vivo* studies of VEGF levels in serum have not been longitudinal or looked at more than 50 patients. It will only be by monitoring large numbers of patients, over time, that the relationship between VEGF levels in serum and factors such as CD4 count, viral load, KS or other vasculopathies, will be determined. Besides CD4 count and viral load measurements there are very few prognostic markers of clinical progression and the development of AIDS-associated diseases. The use of microarrays to identify other core response factors, up regulated and secreted during infection, could significantly reduce the amount of time spent searching for new biomarkers in the future. This study, for example, also showed up regulation of TIMP-2. As described in section 4.3.2.5, Tat cooperates with certain factors in inducing MMP activation and increasing cell membrane-associated TIMP-2, but decreasing the amount of secreted TIMP-1 and –2 (Toschi *et al.,* 2001). These *in vitro* effects are associated with the induction of vascular permeability and oedema *in vivo*, resulting in AIDS-related vasculopathy. As increased levels of MMP-2 are found in plasma from patients with AIDS-KS, compared with HIV-1-uninfected individuals with classic-KS (Toschi *et al.,* 2001), these results suggest that MMP-2 may also be a useful biomarker for HIV-related vasculopathies, or indeed be a therapeutic target to reduce KS aggressiveness or vasculopathy in HIV positive individuals. Similarly, changes in plasma TIMP-2 may of prognostic or diagnostic use.

In addition to the identification of potential biomarkers of HIV disease, a better understanding of the core response to HIV infection also enables the identification of potential novel therapeutic targets. There has long been a focus on targeting and disabling the gene products of HIV in an attempt to prevent viral replication. Increasing attention, however, is now being paid to the host and how host genes may be targeted to prevent successful viral replication. A recent study of the role of host factors in HIV transcription, for example, offers a strong precedent for exploring knockdown of CDK9/CyclinT1 (P-TEFb) and other cellular proteins as a potential new strategy for the development of AIDS therapeutics (Chiu *et al.,* 2004). As described, this study has identified several potential candidates for such targeted inhibition, including VEGF, FOS and TIP30 (sections 4.4.2.1-3). These are drawn from the list of core up regulated genes, which includes several others.

Figure 4.25 Genes regulated by SREBPs.
The diagram shows the major metabolic intermediates in the pathways for synthesis of cholesterol, fatty acids, and triglycerides. *In vivo*, SREBP-2 preferentially activates genes of cholesterol metabolism, whereas SREBP-1c preferentially activates genes of fatty acid and triglyceride metabolism. DHCR, 7-dehydrocholesterol reductase; FPP, farnesyl diphosphate; GPP, geranylgeranyl pyrophosphate synthase; CYP51, lanosterol 14α-demethylase; G6PD, glucose-6-phosphate dehydrogenase; PGDH, 6-phosphogluconate dehydrogenase; GPAT, glycerol-3-phosphate acyltransferase (Reproduced from Horton *et al.*, 2002).



Components of the cholesterol biosynthesis pathway illustrated in Figure 4.25, for example, were strongly up regulated (section 4.3.2.5): acetoacetyl Coenzyme A thiolase; 3-HMG-CoA reductase (HMG-CoA); and mevalonate kinase. It has previously been shown that the use of statin compounds can prevent HIV-1 infection in cultured primary cells, in animal models and in chronically infected individuals (Maziere *et al.*, 1994, Giguère and Tremblay, 2004, del Real *et al.*, 2004). Specifically, statins are HMG-CoA reductase inhibitors normally used to treat hypercholesterolaemia. As well as its role in cholesterol biosynthesis (Figure 4.25), however, HMG-CoA reductase also generates isoprenoids that modify specific cell proteins post-translationally. Rho guanosine triphosphatases (GTPases), for example, must be prenylated at their C terminus by HMG-CoA reductase for their function as molecular switches that cycle between GTP-bound (active) and GDP-bound (inactive) states to control actin cytoskeleton remodelling in response to stimuli (Etienne-Manneville and Hall, 2002). By targeting HMG-CoA, therefore, statins block cholesterol biosynthesis,

but also affect actin cytoskeleton rearrangement by inhibiting Rho GTPases (Koch *et al.,* 1997), required for virus entry and exit. In addition, statins may diminish HIV-1 attachment to target cells by suppressing ICAM-1-LFA-1 interactions (Giguère and Tremblay, 2004). It has also been suggested, however, that the RhoA signal transduction pathway may inhibit HIV-1 replication via a novel effector activity (Wang *et al.,* 2000). This is supported by the recent observation that lovastatin has both pro-HIV (increasing viral genome transcription) and anti-HIV effects (inhibiting virus entry into and virus production by the target cell), presumably by inhibition of Rho activity (del Real *et al.,* 2004). With this in mind, therefore, perhaps more targeted forms of inhibition of factors affected by statins should be explored. If a factor alternative to Rho could be targeted in terms of disrupting cytoskeletal rearrangement, for example, then the Rho activity of inhibiting viral transcription could be left intact. The observed strong up regulation of acetoacetyl Coenzyme A thiolase indicates that this factor is important in HIV infection, so perhaps a targeted inhibition of this gene, rather than one further down the cholesterol biosynthetic pathway, could prove advantageous. It would be interesting to see whether inhibition of acetoacetyl Coenzyme A thiolase would have a similar effect to inhibition of HMG-CoA reductase *in vitro.* If this were the case it would point to an important involvement of increased intracellular cholesterol in HIV replication. Because HMG-CoA reductase inhibition can affect viral replication through several pathways, a more targeted approach will also reveal more about the relationship between HIV and the host cell.

The up regulation of the factors adaptor-related protein complex 2 (AP2) and surfeit 4, involved in intracellular trafficking, was also observed as part of the core response to HIV infection. AP2 is involved in the process of Nef-mediated CD4 down regulation (section 4.3.2.5), but siRNA induced AP2 knockdown has not been shown to have a significant effect on the extent of CD4 down regulation (Rose *et al.,* 2004). These experiments, however, were performed outside of the context of infection so the effect of AP2 knock-down on viral replication was not determined. That is, as this gene is strongly up regulated if may be that AP2 not only functions in mechanisms of protein sorting, resulting in aberrant trafficking, intracellular trapping and degradation of targeted receptors, but also in the trafficking of viral proteins, such as Env. The strong up regulation of surfeit 4 may also bear investigation for the same reasons.

Other potential therapeutic targets for which less information is available include the factors involved in transcription found to be up regulated as part of the core response: AP2 gamma, SERTA domain containing 2, and zinc finger protein 131. The strong up regulation of these genes suggests that they may play an important role in the process of infection, either in aiding viral transcription or the transcription of host genes in response to infection, making them, like P-TEFb, potential targets for the development of AIDS therapeutics (Chiu *et al.*, 2004).

4.4.4 The diverging response to HIV infection

As described in Chapter 3, both epidemiological and *in vitro* studies have provided evidence that sequence variation among the genetic variants of HIV may influence disease progression, transmissibility, and viral phenotype. In terms of *in vitro* studies, investigations have largely focussed on the importance of LTR/accessory gene sequence divergence and the functional consequences of these, but none to date have investigated whether such well characterised sequence divergence influences patterns of host gene expression, during infection. Several microarray-based studies of host gene expression during HIV-1 infection, however, have greatly increased our understanding of how human cells respond to HIV-1 infection and how HIV-1 is capable of modulating host gene expression. These studies have therefore set the scene for this investigation, in which comparison of genetic variants of HIV has not only enabled the characterisation of a core transcriptional response, but has also enabled the identification of examples of divergent gene expression, in which two viruses cause directly the opposite effect on expression of a particular gene. As with the core response genes, these transcriptional responses require confirmation (beyond the scope of this thesis) but the preliminary data presented in this study provide a clear insight into how different genetic variants of HIV may differentially modulate gene expression, especially in relation to control of cell cycle and apoptosis, and transcriptional regulation. Taken together, these results provide some suggestions with regards to how the genetic variation between HIV-1 subtype B, group O and HIV-2 affects viral phenotype *in vitro,* and therefore indirectly how it may influence the process of infection and disease *in vivo.*

4.4.4.1 Transcriptional divergence detected relates to published observed differences between HIV-1 and HIV-2

As described previously it has long been known that HIV-2 disease progression seems to be much slower than HIV-1, and that HIV-2 is associated with decreased transmissibility, protracted stage of CD4 decline and significantly lower viral load (literature reviewed in Chapter 3, section 3.1.1.5). This has, in part, been attributed to certain *in vitro* observations, which are listed here. The data from this study in support of these is summarised.

HIV-2 infection appears to result in lower viral replication compared to HIV-1 (Berry et al., 1998, Popper et al., 2000).

Whilst HIV-2 specific immune responses are more broad and efficient than those detected in HIV-1 infected individuals, potentially explaining this observation (reviewed in Andersson, 2001), other factors are likely to contribute. These may include the differential expression of host factors involved in viral transcription, such as TFIIH, which interacts with HIV-1 Tat as a component of the HIV-1 transcription pre-initiation complex (Garcia-Martinez *et al.,* 1997). Expression of this gene was down regulated in response to HIV-2, but was up regulated in response to HIV-1 subtype B. Together with up regulation of cyclin dependent kinase 7 (Cdk7 - an essential component of the TFIIH transcription factor) in response to HIV-1 subtype B, these data suggest a difference between HIV-1 and HIV-2 in the expression of host factors essential for viral transcription, where transcription from the HIV-1 LTR is favoured. Differential regulation of GNC5, BCl6 and RXRA may also contribute to the differences between HIV-1 and HIV-2 in terms of transcriptional activity and subsequent viral replication rate.

HIV-2 is associated with a lower general level of immune activation and apoptosis compared with HIV-1 (Machuca et al., 2004)

Up regulation of Pim-2 oncogene in response to O and ETP infection, but down regulated during subtype B infection, provides a direct link between HIV-2 and reduced apoptosis (DeBiasi *et al.,* 2003). Confirmation of this observation will prove important in understanding this potentially crucial difference between HIV-1 and HIV-2.

HIV-2 infection is associated with increased CCR5 down regulation (Shea et al., 2004). LAG1 was up regulated during HIV-2 ETP infection, but was down regulated during HIV-1 group O infection. This may be important as LAG1 expression affects ceramide synthesis which, in addition to effecting growth, proliferation, stress resistance, and apoptosis (Jazwinski and Conzelmann, 2002) also perturbs localised membrane domain structure and organization (Finnegan *et al.*, 2004). Infectivity may be reduced as a result of this, due to trafficking of virions via an endocytic pathway that leads to non-productive infection (Schaeffer *et al.*, 2004, Finnegan *et al.*, 2004) and disruption of coreceptor recycling. Hence both viral infectivity and CCR5 expression may be abrogated as a result of LAG-1 up regulation, during HIV-2 infection. It has been shown that by enhancing ceramide levels in $CD4^+$ lymphocytes and in monocyte-derived macrophages with 4-HPR or Smase, HIV-1 infectivity can be significantly reduced without toxicity, making these drugs potentially suitable as anti-HIV therapeutics (Finnegan *et al.*, 2004).

4.4.4.2 HIV-1 group O also differentially regulates host cell transcription

The majority of this discussion has focussed on the transcriptional response of host cells to HIV-1 versus HIV-2 infection, as the difference between these two viruses has been fairly well studied both *in vitro* and *in vivo*. HIV-1 group O infections, perhaps due to their rarity, are much less well studied and very few publications have considered the importance of their genetic distance from the HIV-1 group M subtypes. Despite the lack of corroborating data to support observations, as with the HIV-2 microarray data, several interesting findings were made.

As in HIV-2 infection, Pim2 was up regulated in response to group O infection, inviting the possibility that group O infection may also be associated with decreased apoptosis *in vitro* and potentially, therefore, an altered pathogenicity *in vivo* compared to HIV-1 subtype B. Up regulation of RXRA was also detected in group O infection, whereas down regulation of this nuclear receptor was observed in subtype B infection. As antagonists of this receptor activate HIV-1 replication it is possible that modulation of the retinoid receptor may have an important effect on regulation of HIV-1 group O infection. Whether certain subtypes or types of HIV differ in their response to RXRA-antagonists, perhaps through variation in their LTR sequence and thus responsiveness to repression by host factors, is not known. Finally, cdk5 is actively up regulated in

response to group O infection, but was down regulated in response to subtype B. Whilst the biological importance of this in T-cells is not particularly clear, it is likely that this divergent regulation is indicative of the different way in which HIV-1 subtype B and group O Tat can act on host cells.

4.4.5 Summary

In considering both the core and diverging responses to HIV infection identified in this study, one must conclude that they represent the combination of both cell- and viral-driven changes in gene expression that are the most strong and essential, 24 hours post-infection of ST1-R5 T-cells. Not all processes of transcriptional modulation taking place within the infected cell, however, will have been described.

In terms of the core response, given the sequence divergence between HIV-1 B, O and HIV-2, it is tempting to consider that the host's part in this core infection profile is perhaps more dominant than in the diverging responses identified: the feature that unifies each infection condition is the host, after all. Evidence has been presented, however, that points to a strong viral involvement in the modulation of this core response (section 4.3.2.5) and of the genes described the regulation of many, at least in part, can be attributed to the actions of Tat and Nef.

Evidence has also been presented that shows, in addition to the core response to HIV infection, different genetic variants of HIV induce divergent gene expression patterns. In certain cases such as Pim2 regulation, these lend support to published observations, in particular in relation to the differences between HIV-1 and HIV-2. Other observations point to as yet uncharacterised differences between the way HIV-1 B, group O and HIV-2 interact with the host, and the significance of these is consequently less clear. In understanding the differences between different genetic variants of HIV and their interactions with the host, we may begin to better understand the significance of viral genetic variation on HIV disease. The comparison of HIV-1 and HIV-2 is very important in this respect. Between the different HIV subtypes, however, the effects of genetic variation are likely to be subtle, and significantly complicated by the influences of host genetic variation and other environmental factors. The importance of subtype variation in the development of drug resistance may be more important. The core host and viral responses to infection, however, are the unifying factor between HIV infection

and the fact that infection will ultimately lead to progressive CD4 decline, AIDS, opportunistic infections and death. By characterising and better understanding the most conserved consequences of HIV infection a clearer view of the interaction between host and virus, and how it may be interrupted therapeutically, may be sought.

Chapter 5.0 Discussion and future work

In the introduction to this thesis the aim was set out as being to characterise the genotypic and phenotypic features of the different HIV-1 types and subtypes, the question being asked are the genotypic differences between HIV variants manifested phenotypically and if so, does it matter?

In Chapter 2, the development of a HIV subtyping tool, STAR, was described. This was done bearing in mind that, globally, laboratory monitoring of HIV-1 drug resistance has the potential to provide an excellent sequence resource for identifying HIV subtypes, their movement and by association with clinical data, their clinical significance. Any clinical laboratory dealing with drug resistance profiling of potentially thousands of sequences but where subtyping is not the primary concern, however, will require an appropriate method for such analysis. The genotypic variation between the HIV-1 subtypes in Pol was therefore thoroughly characterised, to form the basis of a PSSM-based subtyping algorithm. The resulting program, STAR, will automatically and accurately classify 36 sequences per minute and is independent of the user. It does not attempt to define subtype boundaries, or replace the phylogenetics and phylogenies from which it was derived; rather it is designed as an aid to epidemiological investigations and the study of the importance of HIV-1 genetic variation. This will only be achieved, however, if STAR is implemented in standardized settings and a concerted effort is made in the processes of acquisition, compilation and annotation of HIV sequence data produced for resistance testing. Current efforts to achieve this are ongoing, in the creation of HIVCentral; a database for resistance sequence data and associated clinical information produced across Europe (R. Gifford, UCL). A recently updated version of STAR has been implemented in this database as the main method for subtyping sequence data (Myers *et al.*, in press), so it is hoped that in this setting questions about the relationship between HIV subtype and: disease progression; efficiency of diagnostics; response to therapy; drug resistance patterns; and the introduction and spread of genetically distinct HIV variants into monitored populations, may be asked. Given the ability of STAR to detect recombinants within Pol, such a large-scale analysis may also enable an estimation of recombination frequencies within this genomic region. It will only be with such large

numbers of sequences and accurate subtype classification that, at the population level, we may begin to see the importance of HIV-1 genotype.

The need to develop STAR for such studies therefore naturally alludes to the fact that the role viral genotype plays in the phenotype of HIV disease is, despite several studies on the matter, far from clear. Many would argue that the subtle differences so far detected are more easily attributable to host or environmental variation, moulding the process of disease. A cogent argument for the importance of founder effects, together with a paucity of data with regards to inter-host evolution of HIV in different host populations, leaves the notion that biological differences between HIV subtypes might exist by virtue of their evolution in genetically distinct host environments, with little substance. The importance of intra-host evolution in relation to disease progression, however, is well documented; so the question remains in this context do the genetic differences between different HIV subtypes modulate intra-host evolution and therefore progression of disease? As the virus is intrinsically part of the host, its genetic composition moulded by the environment in which it exists, it seems likely that on a person to person basis if you start with a different viral genetic variant then this interaction may not always have the same end result. The data presented in Chapter 3 of this thesis display how, when HIV of different subtypes are placed in the same experimental system, they do indeed behave differently. Whilst this may not be a direct reflection of the *in vivo* situation, as a means to outline the potential for phenotypic differences between genetic variants of HIV and the extreme difficulty in characterising them, these experiments are very clear. In addition, whilst drawing specific conclusions from these data is difficult, it does seem that *env* tropism and coreceptor usage are very important in determining viral phenotype *in vitro*. The mechanisms of this have been discussed and are less than clear, but if the genetic variation between HIV-1 subtypes does guide their coreceptor preference, as preliminary data presented here suggests it might, then it seems inevitable that the HIV-1 subtypes and the subtleties of the process of disease they cause, may differ.

In consideration of the experiments required to determine the importance of this relationship between subtype, coreceptor usage and disease progression, one must first consider that the reasons why 50% of viral isolates from patients with AIDS display X4-tropism, but the presence of X4-tropic virus in HIV+ patients at seroconversion is

rare, are not completely understood. Review of the literature, however, enables a general process to be described in which a selective bottleneck at transmission results in positive selection of R5-tropic viruses, which predominate due to their preferential access to and replication within body compartments favouring virus of this tropism, such as GALT. Over time this becomes exhausted and viruses of X4 tropism, which may have been present as minority quasispecies within CXCR4+ cellular compartments, gain a competitive advantage (Grossman and Paul, 2000, Moore *et al.,* 2004). The chronic activation of CD4+ T-cells has, however, begun the decline in their functionality and concomitant with switching and outgrowth of X4-tropic virus, a precipitous drop in CD4+CXCR4+ T-cell count occurs, the broad infection capacity of X4-tropic viruses (targets including haemopoeitic progenitor cells and thymocytes, Bleul *et al.,* 1997, Berkowitz *et al.,* 1998, Rosu-Myles *et al.,* 2000) enhancing immune dysfunction. In combination with destruction of CCR5+ compartments, general immune dysfunction due to chronic and aberrant activation, and loss of CD4+ T-cells, AIDS is the result. In this setting the important consideration, therefore, is the time at which R5 to X4 switch occurs, the suggestion being that the sooner you switch the worse your prognosis. It is therefore of interest whether different viral subtypes will switch at different times, by virtue of their genetic differences within gp120 placing them at different points on the evolutionary continuum from R5 to X4-usage. In order to determine this, longitudinal studies will be required which follow patients of matched ethnicity, age, sex, and preferably treatment, but which are infected with different HIV subtypes. Whilst in the past such patient cohorts have been few and far between, clinical trials of antiretroviral therapies, especially those that are beginning to take place in Africa where multiple subtypes of HIV co-circulate, may provide populations appropriate for such studies. Envelope sequencing over time would be realistic in this setting, and may indicate the prognostic value of V3 scoring (Jensen and van 't Wout, 2003). If there is an association between changes in V3 score, tropism, viral subtype, and the phenotype of HIV disease, it would hopefully be detected. It would also be advantageous to repeat experiments performed in this thesis, only with a much larger sample of viruses representative of each subtype. It would be interesting to see if patterns of coreceptor preference among HIV primary isolates of different subtype are borne out, and whether these relate to growth phenotype *in vitro,* and perhaps therefore also the phenotype of disease in the patient from which they were acquired.

Despite the concentration on the importance of coreceptor usage and its relationship with HIV subtype and disease, however, this thesis by no means concludes that this is the all-important manifestation of HIV genetic variation that may modulate the progression of infection. One only needs to consider HIV-2 as an example of a virus with a typically very broad coreceptor usage and therefore cellular tropism, but which has a very attenuated clinical course. Since HIV-2 varies from HIV-1 in approximately 50% of its genome sequence and between the HIV-1 subtypes this variation may be up to 30%, the questions follow; other than coreceptor usage what other features of HIV govern its interaction with the cell; how do these interactions differ, between genetic variants; and do they effect the phenotype of disease resulting? Clearly these are the all important questions, all of which cannot be addressed at once, if at all. Accordingly, the fourth chapter of this thesis describes a series of microarray experiments which aimed to answer the question: to what degree is there commonality in the way genetically divergent HIV isolates interact with the host cell, in terms of their modulation or induction of host transcriptional processes? In essence, unlike the previous chapter, this work was concerned with understanding the most specific and intimate interactions between HIV and its host cell. With the resulting information it was hoped that a better approach might be sought in dissecting the roles host and virus play in the process of infection and, furthermore, whether genetic variation of HIV is manifested in these processes.

In terms of achieving a better understanding of the interaction between host and virus during infection, the results of chapter four largely confirmed that which has previously been described: HIV infection of T-cells generally results in reduced cell cycling and up regulation of certain genes required for viral replication and spread (generally accepted to be as a result of expression of certain viral accessory genes, in particular Tat) and the initiation of a cellular stress-type response. Although not confirmed by additional experiments or in other publications, certain more directed mechanisms of host anti-viral defence and ways in which the virus may modulate cellular transcription processes were also described. The strength in this study, however, lies in the fact that three genetically distinct HIV isolates were studied in the same system: HIV-1 subtype B, HIV-1 group O and HIV-2 ETP. That is, in comparing gene expression in response to these viruses, an extra dimension could be added to the results obtained in that patterns of gene expression could be segregated, firstly into those that are 'core responses',

which whilst unlikely to be a complete description, may be taken to represent examples of the most vital and conserved processes of genetic regulation, orchestrated by both host and virus, during HIV infection. An analysis of the core response was performed and the fact that certain genes identified may be novel drug targets or biomarkers of disease further emphasises the utility of this sort of study. Secondly 'diverging responses', seen in response to individual viruses, were identified. Careful analysis of data and curation of such gene lists further enabled the distinction between genes which were uniquely regulated between viruses, and those for which two viruses caused directly the opposite effect on expression. In asking what the significance of this is in the absence of complementary phenotypic data, the study of HIV-1 and HIV-2 proved advantageous as possible differences in gene expression detected may go some way to explaining certain well-defined differences between these viruses in terms of the way they behave *in vitro*. The differential modulation of apoptosis, for example, illustrates how changes at the transcriptional level may be manifested *in vitro,* and consequently also *in vivo*. With these data in hand, therefore, it is possible to answer the question posed, in that there is a commonality in the way in which these three genetically distinct HIV isolates interact with T-cells, as one might expect. There are also, however, diverging responses to the different viruses implying that genetic variation within HIV is important in modulating host gene expression. This study therefore sets a precedent for further investigation as to whether the genetic variation between the HIV-1 subtypes will be manifested, as between HIV-1 and HIV-2, as differences in modulation of host gene expression that are reflected *in vitro,* and which have implications *in vivo*. It will be interesting to see if, as with differences in coreceptor usage brought about by sequence variation, whether the different HIV-1 subtypes also lie on a continuum of differential transcriptional modulation between HIV-1 subtype B, group O and HIV-2, depending on their genetic distance from each.

The fourth chapter of this thesis is incomplete in that it lacks confirmation of gene expression data. Ideally, changes in gene expression should not only be confirmed by RT-PCR or quantitative PCR at the RNA level, but levels of expressed protein should be determined. Together with knock out or inhibition of gene products up regulated as part of core or diverging responses, these analysis would lend much greater weight to the biological and functional consequences of observations here made. These were unfortunately beyond the scope of this thesis. As an indication of the methodological

optimisation and analytical rigour required for such a study, however, this work is complete. And as an initial foray into the study of the interaction of HIV primary isolates with the host cell, it perhaps is a positive thing that this work creates more questions than answers.

Thus, in returning to the question asked at the beginning of this thesis it can be concluded that the genotypic differences between the HIV subtypes are manifested phenotypically, when studied *in vitro*: both in the manner in which they grow, largely governed by preferential coreceptor usage; and in the way they modulate host transcription, orchestrated by both the virus' effect on the host and the hosts response to the virus. In considering whether the phenotypic manifestations of genotypic variation matter, therefore, attention should perhaps be paid to the latter point: both the virus and the host determine the outcome of HIV infection. That is, whilst genetic variation amongst HIV almost certainly has some effect on the phenotype of disease, it is perhaps the host that has the greater influence. Despite HIV's far greater capacity for genetic diversification, the consequences of genetic and behavioural diversity in humans are likely to outstrip that which can be achieved by the genetic variation between the HIV-1 subtypes: in analysing data at a population level relating HIV-1 subtype to features of HIV disease, it may be that important features of human host variation will be identified. The essential problem, however, is that all such analyses rely on pre-existing classifications, such as viral subtype, risk group, sex, ethnicity, and so on. It may be that hidden within and across these groupings there are variants of both virus and host in which the phenotype of HIV-induced disease, for one reason or another, is different. It will only be when such people and viral variants begin to be identified that, together with current classifications, we may truly begin to understand the importance of viral and host genetic variation. Recent ground breaking research has indeed added insight to our consideration of the importance of host genetic variation. For the time being, in terms of the importance of viral variation, a greater understanding as to why HIV-2 fails to cause a significant disease burden but HIV-1 continues to newly infect and kill millions of people worldwide every year, would be a good place to start.

Chapter 6.0 References

Abebe A, Pollakis G, Fontanet AL, et al. Identification of a genetic subcluster of HIV type 1 subtype C (C') widespread in Ethiopia. AIDS Res Hum Retroviruses 2000,16:1909-1914.

Adams SG, Dohner DE, Gelb LD. Restriction fragment differences between the genomes of the Oka varicella vaccine virus and American wild-type varicella-zoster virus. J Med Virol 1989,29:38-45.

Adje C, Cheingsong R, Roels TH, et al. High prevalence of genotypic and phenotypic HIV-1 drug-resistant strains among patients receiving antiretroviral therapy in Abidjan, Cote d'Ivoire. J Acquir Immune Defic Syndr 2001,26:501-506.

Adjorlolo-Johnson G, De Cock KM, Ekpini E, et al. Prospective comparison of mother-to-child transmission of HIV-1 and HIV-2 in Abidjan, Ivory Coast. Jama 1994,272:462-466.

Agostini HT, Ryschkewitsch CF, Singer EJ, Stoner GL. JC virus regulatory region rearrangements and genotypes in progressive multifocal leukoencephalopathy: two independent aspects of virus variation. J Gen Virol 1997,78 (Pt 3):659-664.

Agwale SM, Zeh C, Paxinos E, et al. Genotypic and Phenotypic Analyses of Drug Susceptibility in HIV-1 Isolates from Drug-Naive Patients in Nigeria. 9th Conference on Retroviruses and Opportunistic Infections. 24-28-4 Feb 2002, Seattle, WA.

AIDS Activity Center For Infectious Diseases Centers For Disease Control:
Acquired Immunodeficiency Syndrome (AIDS) weekly surveillance report- United States. 1983.

Alaeus A, Lidman K, Bjorkman A, Giesecke J, Albert J. Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. Aids 1999,13:901-907.

Alba MM, Lee D, Pearl FM, et al. VIDA: a virus database system for the organization of animal virus genome open reading frames. Nucleic Acids Res 2001,29:133-136.

Alizon M, Sonigo P, Barre-Sinoussi F, et al. Molecular cloning of lymphadenopathy-associated virus. Nature 1984,312:757-760.

Alkhatib G, Combadiere C, Broder CC, et al. CC CKR5: a RANTES, MIP-1alpha, MIP-1beta receptor as a fusion cofactor for macrophage-tropic HIV-1. Science 1996,272:1955-1958.

Alon U, Barkai N, Notterman DA, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A 1999,96:6745-6750.

Altman. New homosexual disorder worries officials. New York Times 1982.

Ameglio F, Tilocca F, Arca MV, Alemanno L, Dolei A. Ferritin downregulation in HIV-infected cells. AIDS Res Hum Retroviruses 1993,9:795-798.

Amornkul PN, Tansuphasawadikul S, Limpakarnjanarat K, et al. Clinical disease associated with HIV-1 subtype B' and E infection among 2104 patients in Thailand. Aids 1999,13:1963-1969.

Anastassopoulou CG, Kostrikis LG. The impact of human allelic variation on HIV-1 disease. Curr HIV Res 2003,1:185-203.

Anastos K, Kalish LA, Hessol N, et al. The relative value of CD4 cell count and quantitative HIV-1 RNA in predicting survival in HIV-1-infected women: results of the women's interagency HIV study. Aids 1999,13:1717-1726.

Andersson S. HIV-2 and the Immune Response. AIDS Reviews 2001,3:11-23.

Ando T, Jin Q, Gentsch JR, et al. Epidemiologic applications of novel molecular methods to detect and differentiate small round structured viruses (Norwalk-like viruses). J Med Virol 1995,47:145-152.

Andreasson PA, Dias F, Naucler A, Andersson S, Biberfeld G. A prospective study of vertical transmission of HIV-2 in Bissau, Guinea-Bissau. Aids 1993,7:989-993.

Andrieu JM, Even P, Venet A, et al. Effects of cyclosporin on T-cell subsets in human immunodeficiency virus disease. Clin Immunol Immunopathol 1988,47:181-198.

Andrieu M, Chassin D, Desoutter JF, et al. Downregulation of major histocompatibility class I on human dendritic cells by HIV Nef impairs antigen presentation to HIV-specific CD8+ T lymphocytes. AIDS Res Hum Retroviruses 2001,17:1365-1370.

Aoki Y, Tosato G. Targeted inhibition of angiogenic factors in AIDS-related disorders. Curr Drug Targets Infect Disord 2003,3:115-128.

Apcher GS, Heink S, Zantopf D, et al. Human immunodeficiency virus-1 Tat protein interacts with distinct proteasomal alpha and beta subunits. FEBS Lett 2003,553:200-204.

Apetrei C, Descamps D, Collin G, et al. Human immunodeficiency virus type 1 subtype F reverse transcriptase sequence and drug susceptibility. J Virol 1998,72:3534-3538.

Apetrei C, Loussert-Ajaka I, Descamps D, et al. Lack of screening test sensitivity during HIV-1 non-subtype B seroconversions. Aids 1996,10:F57-60.

Arendt CW, Littman DR. HIV: master of the host cell. Genome Biol 2001,2:REVIEWS1030.

Arens M. Methods for subtyping and molecular comparison of human viral genomes. Clin Microbiol Rev 1999,12:612-626.

Arens M, Swierkosz EM. Detection of rotavirus by hybridization with a nonradioactive synthetic DNA probe and comparison with commercial enzyme immunoassays and silver-stained polyacrylamide gels. J Clin Microbiol 1989,27:1277-1279.

Arts GJ, Kuersten S, Romby P, Ehresmann B, Mattaj IW. The role of exportin-t in selective nuclear export of mature tRNAs. Embo J 1998,17:7430-7441.

Ascherl G, Hohenadl C, Schatz O, et al. Infection with human immunodeficiency virus-1 increases expression of vascular endothelial cell growth factor in T cells: implications for acquired immunodeficiency syndrome-associated vasculopathy. Blood 1999,93:4232-4241.

Ball SC, Abraha A, Collins KR, et al. Comparing the ex vivo fitness of CCR5-tropic human immunodeficiency virus type 1 isolates of subtypes B and C. J Virol 2003,77:1021-1038.

Barat C, Gervais P, Tremblay MJ. Engagement of ICAM-3 Provides a Costimulatory Signal for Human Immunodeficiency Virus Type 1 Replication in both Activated and Quiescent CD4+ T Lymphocytes: Implications for Virus Pathogenesis. J. Virol. 2004,78:6692-6697.

Barlow KL, Tatt ID, Cane PA, Pillay D, Clewley JP. Recombinant strains of HIV type 1 in the United Kingdom. AIDS Res Hum Retroviruses 2001,17:467-474.

Baron BW, Anastasi J, Thirman MJ, et al. The human programmed cell death-2 (PDCD2) gene is a target of BCL6 repression: implications for a role of BCL6 in the down-regulation of apoptosis. Proc Natl Acad Sci USA 2002,99:2860-2865.

Baron BW, Desai M, Baber LJ, et al. BCL6 can repress transcription from the human immunodeficiency virus type I promoter/enhancer region. Genes Chromosomes Cancer 1997,19:14-21.

Barre-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science 1983, 220:868-871.

Bassett DE, Jr., Eisen MB, Boguski MS. Gene expression informatics--it's all in your mine. Nat Genet 1999,21:51-55.

Belden WJ, Barlowe C. Role of Erv29p in collecting soluble secretory proteins into ER-derived transport vesicles. Science 2001,294:1528-1531.

Ben-Dor A, Shamir R, Yakhini Z. Clustering gene expression patterns. J Comput Biol 1999,6:281-297.

Benelli R, Barbero A, Ferrini S, et al. Human immunodeficiency virus transactivator protein (Tat) stimulates chemotaxis, calcium mobilization, and activation of human polymorphonuclear leukocytes: implications for Tat-mediated pathogenesis. J Infect Dis 2000,182:1643-1651.

Bennin DA, Don AS, Brake T, et al. Cyclin G2 associates with protein phosphatase 2A catalytic and regulatory B' subunits in active complexes and induces nuclear aberrations and a G1/S phase cell cycle arrest. J Biol Chem 2002,277:27449-27467.

Beraud F, Kessler N, Aymard M. Contribution of monoclonal antibodies to the study of parainfluenza virus antigens. Dev Biol Stand 1984,57:257-267.

Berkowitz RD, Beckerman KP, Schall TJ, McCune JM. CXCR4 and CCR5 expression delineates targets for HIV-1 disruption of T cell differentiation. J Immunol 1998,161:3702-3710.

Berry N, Ariyoshi K, Jaffar S, et al. Low peripheral blood viral HIV-2 RNA in individuals with high CD4 percentage differentiates HIV-2 from HIV-1 infection. J Hum Virol 1998,1:457-468.

Bhoopat L, Eiangleng L, Rugpao S, et al. In vivo identification of Langerhans and related dendritic cells infected with HIV-1 subtype E in vaginal mucosa of asymptomatic patients. Mod Pathol 2001,14:1263-1269.

Biasin MR, Fiordalisi G, Zanella I, Cavicchini A, Marchelle G, Infantolino D. A DNA hybridization method for typing hepatitis C virus genotype 2c. J Virol Methods 1997,65:307-315.

Bigger CB, Brasky KM, Lanford RE. DNA microarray analysis of chimpanzee liver during acute resolving hepatitis C virus infection. J Virol 2001,75:7059-7066.

Birnboim HC, Doly J. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res 1979,7:1513-1523.

Blaak H, van't Wout AB, Brouwer M, Hooibrink B, Hovenkamp E, Schuitemaker H. In vivo HIV-1 infection of CD45RA(+)CD4(+) T cells is established primarily by syncytium-inducing variants and correlates with the rate of CD4(+) T cell decline. Proc Natl Acad Sci U S A 2000,97:1269-1274.

Blackard JT, Renjifo B, Fawzi W, et al. HIV-1 LTR subtype and perinatal transmission. Virology 2001,287:261-265.

Bleul CC, Wu L, Hoxie JA, Springer TA, Mackay CR. The HIV coreceptors CXCR4 and CCR5 are differentially expressed and regulated on human T lymphocytes. Proc Natl Acad Sci U S A 1997,94:1925-1930.

Bobat R, Moodley D, Coutsoudis A, Coovadia H. Breastfeeding by HIV-1-infected women and outcome in their infants: a cohort study from Durban, South Africa. Aids 1997,11:1627-1633.

Bolt G, Berg K, Blixenkrone-Moller M. Measles virus-induced modulation of host-cell gene expression. J Gen Virol 2002,83:1157-1165.

Bouyac-Bertoia M, Dvorin JD, Fouchier RA, et al. HIV-1 infection requires a functional integrase NLS. Mol Cell 2001,7:1025-1035.

Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet 2001,29:365-371.

Brennan RO, Durack DT. Gay compromise syndrome. Lancet 1981,2:1338-1339.

Bres V, Tagami H, Peloponese JM, et al. Differential acetylation of Tat coordinates its interaction with the co-activators cyclin T1 and PCAF. Embo J 2002,21:6811-6819.

Briggs JA, Wilk T, Welker R, Krausslich HG, Fuller SD. Structural organization of authentic, mature HIV-1 virions and cores. Embo J 2003,22:1707-1715.

Brindeiro PA, Brindeiro RM, Mortensen C, et al. Testing genotypic and phenotypic resistance in human immunodeficiency virus type 1 isolates of clade B and other clades from children failing antiretroviral therapy. J Clin Microbiol 2002,40:4512-4519.

Bron R, Klasse PJ, Wilkinson D, et al. Promiscuous use of CC and CXC chemokine receptors in cell-to-cell fusion mediated by a human immunodeficiency virus type 2 envelope protein. J Virol 1997,71:8405-8415.

Brosch R, Gordon SV, Pym A, Eiglmeier K, Garnier T, Cole ST. Comparative genomics of the mycobacteria. Int J Med Microbiol 2000,290:143-152.

Brown AJ, Richman DD. HIV-1: gambling on the evolution of drug resistance? Nat Med 1997,3:268-271.

Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat Genet 1999,21:33-37.

Browne EP, Wing B, Coleman D, Shenk T. Altered cellular mRNA levels in human cytomegalovirus-infected fibroblasts: viral block to the accumulation of antiviral mRNAs. J Virol 2001,75:12319-12330.

Bruisten SM, Frissen PH, Van Swieten P, et al. Prospective longitudinal analysis of viral load and surrogate markers in relation to clinical progression in HIV type 1-infected persons. AIDS Res Hum Retroviruses 1997,13:327-335.

Buchman TG, Roizman B, Adams G, Stover BH. Restriction endonuclease fingerprinting of herpes simplex virus DNA: a novel epidemiological tool applied to a nosocomial outbreak. J Infect Dis 1978,138:488-498.

Buchman TG, Roizman B, Nahmias AJ. Demonstration of exogenous genital reinfection with herpes simplex virus type 2 by restriction endonuclease fingerprinting of viral DNA. J Infect Dis 1979,140:295-304.

Bukrinsky MI, Sharova N, Dempsey MP, et al. Active nuclear import of human immunodeficiency virus type 1 preintegration complexes. Proc Natl Acad Sci U S A 1992,89:6580-6584.

Bureau of Hygiene & Tropical Diseases: 'AIDS newsletter' Issue. 1986 January 15.

Butera ST, Perez VL, Besansky NJ, et al. Extrachromosomal human immunodeficiency virus type-1 DNA can initiate a spreading infection of HL-60 cells. J Cell Biochem 1991,45:366-373.

Butterfield LH, Merino A, Golub SH, Shau H. From cytoprotection to tumor suppression: the multifactorial role of peroxiredoxins. Antioxid Redox Signal 1999,1:385-402.

Campbell SM, Crowe SM, Mak J. Lipid rafts and HIV-1: from viral entry to assembly of progeny virions. J Clin Virol 2001,22:217-227.

Cane PA, de Ruiter A, Rice P, Wiselka M, Fox R, Pillay D. Resistance-associated mutations in the human immunodeficiency virus type 1 subtype c protease gene from treated and untreated patients in the United Kingdom. J Clin Microbiol 2001,39:2652-2654.

Caplan S, Hartnell LM, Aguilar RC, Naslavsky N, Bonifacino JS. Human Vam6p promotes lysosome clustering and fusion in vivo. J Cell Biol 2001,154:109-122.

Cara A, Cereseto A, Lori F, Reitz MS, Jr. HIV-1 protein expression from synthetic circles of DNA mimicking the extrachromosomal forms of viral DNA. J Biol Chem 1996,271:5393-5397.

Caride E, Brindeiro R, Hertogs K, et al. Drug-resistant reverse transcriptase genotyping and phenotyping of B and non-B subtypes (F and A) of human immunodeficiency virus type I found in Brazilian patients failing HAART. Virology 2000,275:107-115.

Caride E, Hertogs K, Larder B, et al. Genotypic and phenotypic evidence of different drug-resistance mutation patterns between B and non-B subtype isolates of human immunodeficiency virus type 1 found in Brazilian patients failing HAART. Virus Genes 2001,23:193-202.

Carr JK, Avila M, Gomez Carrillo M, et al. Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. Aids 2001,15:F41-47.

Carr JK, Salminen MO, Koch C, et al. Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand. J Virol 1996,70:5935-5943.

Carter KL, Cahir-McFarland E, Kieff E. Epstein-barr virus-induced changes in B-lymphocyte gene expression. J Virol 2002,76:10427-10436.

Carvalho M, Kirkland M, Derse D. Protein interactions with DNA elements in variant equine infectious anemia virus enhancers and their impact on transcriptional activity. J Virol 1993,67:6586-6595.

Casareale D, Dewhurst S, Sonnabend J, Sinangil F, Purtilo DT, Volsky DJ. Prevalence of AIDS-associated retrovirus and antibodies among male homosexuals at risk for AIDS in Greenwich Village. AIDS Res 1984,1:407-421.

Castelain S, Khorsi H, Zawadzki P, et al. Direct blotting electrophoresis for sequencing and genotyping hepatitis C virus. J Virol Methods 1997,65:237-243.

Cavaleiro R, Sousa AE, Loureiro A, Victorino RM. Marked immunosuppressive effects of the HIV-2 envelope protein in spite of the lower HIV-2 pathogenicity. Aids 2000,14:2679-2686.

CDC. Kaposi's Sarcoma (KS), Pneumocystis Carinii Pneumonia (PCP), and Other Opportunistic Infections (01): Cases Reported to CDC as of July 8'. 1982.

Chambers J, Angulo A, Amaratunga D, et al. DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. J Virol 1999,73:5757-5766.

Chang YE, Laimins LA. Microarray analysis identifies interferon-inducible genes and Stat-1 as major transcriptional targets of human papillomavirus type 31. J Virol 2000,74:4174-4182.

Cheingsong-Popov R, Weiss RA, Dalgleish A, et al. Prevalence of antibody to human T-lymphotropic virus type III in AIDS and AIDS-risk patients in Britain. Lancet 1984,2:477-480.

Chen L, Glover JN, Hogan PG, Rao A, Harrison SC. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. Nature 1998,392:42-48.

Cheng-Mayer C, Quiroga M, Tung JW, Dina D, Levy JA. Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. J Virol 1990,64:4390-4398.

Cheng-Mayer C, Seto D, Tateno M, Levy JA. Biologic features of HIV-1 that correlate with virulence in the host. Science 1988,240:80-82.

Chern KC, Chandler DB, Martin DF, Kuppermann BD, Wolitz RA, Margolis TP. Glycoprotein B subtyping of cytomegalovirus (CMV) in the vitreous of patients with AIDS and CMV retinitis. J Infect Dis 1998,178:1149-1153.

Chiu YL, Cao H, Jacque JM, Stevenson M, Rana TM. Inhibition of human immunodeficiency virus type 1 replication by RNA interference directed against human transcription elongation factor P-TEFb (CDK9/CyclinT1). J Virol 2004,78:2517-2529.

Chou LL, Epstein J, Cassol SA, West DM, He W, Firth JD. Oral mucosal Langerhans' cells as target, effector and vector in HIV infection. J Oral Pathol Med 2000,29:394-402.

Chun TW, Justement JS, Lempicki RA, et al. Gene expression and viral prodution in latently infected, resting CD4+ T cells in viremic versus aviremic HIV-infected individuals. Proc Natl Acad Sci U S A 2003,100:1908-1913.

Cimarelli A, Darlix JL. Assembling the human immunodeficiency virus type 1. Cell Mol Life Sci 2002,59:1166-1184.

Clapham PR, McKnight A. Cell surface receptors, virus entry and tropism of primate lentiviruses. J Gen Virol 2002,83:1809-1829.

Clapham PR, McKnight A, Weiss RA. Human immunodeficiency virus type 2 infection and fusion of CD4-negative human cell lines: induction and enhancement by soluble CD4. J Virol 1992,66:3531-3537.

Clarke P, Tyler KL. Reovirus-induced apoptosis: A minireview. Apoptosis 2003,8:141-150.

Clavel F, Guetard D, Brun-Vezinet F, et al. - a Isolation of a new human retrovirus from West African patients with AIDS. Science 1986,233:343-346.

Clavel F, Guyader M, Guetard D, Salle M, Montagnier L, Alizon M. – b Molecular cloning and polymorphism of the human immune deficiency virus type 2. Nature 1986,324:691-695.

Cohen B, Skiena S. - a Efficient split synthesis for targeted libraries. J Comb Chem 2000,2:10-18.

Cohen GB, Rangan VS, Chen BK, Smith S, Baltimore D. - b The human thioesterase II protein binds to a site on HIV-1 Nef critical for CD4 down-regulation. J Biol Chem 2000,275:23097-23105.

Cohen P, Bouaboula M, Bellis M, et al. - c Monitoring cellular responses to Listeria monocytogenes with oligonucleotide arrays. J Biol Chem 2000,275:11181-11190.

Cohrs RJ, Hurley MP, Gilden DH. Array analysis of viral gene transcription during lytic infection of cells in tissue culture with Varicella-Zoster virus. J Virol 2003,77:11718-11732.

Col E, Caron C, Seigneurin-Berny D, Gracia J, Favier A, Khochbin S. The histone acetyltransferase, hGCN5, interacts with and acetylates the HIV transactivator, Tat. J Biol Chem 2001,276:28179-28184.

Connor RI, Sheridan KE, Ceradini D, Choe S, Landau NR. Change in coreceptor use coreceptor use correlates with disease progression in HIV-1--infected individuals. J Exp Med 1997,185:621-628.

Cooney MK, Fox JP, Kenny GE. Antigenic groupings of 90 rhinovirus serotypes. Infect Immun 1982,37:642-647.

Cooper DA, Gold J, Maclean P, et al. Acute AIDS retrovirus infection. Definition of a clinical illness associated with seroconversion. Lancet 1985,1:537-540.

Copeland KF, Heeney JL. T helper cell activation and human retroviral pathogenesis. Microbiol Rev 1996,60:722-742.

Corbeil J, Sheeter D, Genini D, et al. Temporal gene regulation during HIV-1 infection of human CD4+ T cells. Genome Res 2001,11:1198-1204.

Cornali E, Zietz C, Benelli R, et al. Vascular endothelial growth factor regulates angiogenesis and vascular permeability in Kaposi's sarcoma. Am J Pathol 1996,149:1851-1869.

Cornelissen M, van Den Burg R, Zorgdrager F, Goudsmit J. Spread of distinct human immunodeficiency virus type 1 AG recombinant lineages in Africa. J Gen Virol 2000,81:515-523.

Costa LJ, Zheng Y-H, Sabotic J, Mak J, Fackler OT, Peterlin BM. Nef Binds p6* in GagPol during Replication of Human Immunodeficiency Virus Type 1. J. Virol. 2004,78:5311-5323.

Cotter PA, Miller JF. In vivo and ex vivo regulation of bacterial virulence gene expression. Curr Opin Microbiol 1998,1:17-26.

Craig HM, Pandori MW, Guatelli JC. Interaction of HIV-1 Nef with the cellular dileucine-based sorting pathway is required for CD4 down-regulation and optimal viral infectivity. Proc Natl Acad Sci U S A 1998,95:11229-11234.

Cristina J, Moya A, Arbiza J, et al. Evolution of the G and P genes of human respiratory syncytial virus (subgroup A) studied by the RNase A mismatch cleavage method. Virology 1991,184:210-218.

Cron RQ, Bartz SR, Clausell A, Bort SJ, Klebanoff SJ, Lewis DB. NFAT1 enhances HIV-1 gene expression in primary human CD4 T cells. Clin Immunol 2000,94:179-191.

Cuadras MA, Feigelstock DA, An S, Greenberg HB. Gene expression pattern in Caco-2 cells following rotavirus infection. J Virol 2002,76:4467-4482.

Daelemans D, De Clercq E, Vandamme AM. A quantitative GFP-based bioassay for the detection of HIV-1 Tat transactivation inhibitors. J Virol Methods 2001,96:183-188.

Dalgleish AG, Beverley PC, Clapham PR, Crawford DH, Greaves MF, Weiss RA. The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. Nature 1984,312:763-767.

Darwin C. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. 1859. http://www.literature.org/authors/darwin-charles/the-origin-of-species/

Datta P, Embree JE, Kreiss JK, et al. Mother-to-child transmission of human immunodeficiency virus type 1: report from the Nairobi Study. J Infect Dis 1994,170:1134-1140.

de Cock KM, Adjorlolo G, Ekpini E, et al. Epidemiology and transmission of HIV-2. Why there is no HIV-2 pandemic. Jama 1993,270:2083-2086.

de Jong JJ, De Ronde A, Keulen W, Tersmette M, Goudsmit J. - a Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J Virol 1992,66:6777-6780.

de Jong JJ, Goudsmit J, Keulen W, et al. – b Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain differ in syncytium formation and replication capacity. J Virol 1992,66:757-765.

de la Fuente C, Santiago F, Deng L, et al. Gene expression profile of HIV-1 Tat expressing cells: a close interplay between proliferative and differentiation signals. BMC Biochem 2002,3:14.

de la Fuente C, Wang L, Wang D, et al. Paradoxical effects of a stress signal on pro- and anti-apoptotic machinery in HTLV-1 Tax expressing cells. Mol Cell Biochem 2003,245:99-113.

de Noronha CM, Sherman MP, Lin HW, et al. Dynamic disruptions in nuclear envelope architecture and integrity induced by HIV-1 Vpr. Science 2001,294:1105-1108.

de Oliveira T, Engelbrecht S, Janse van Rensburg E, et al. Variability at human immunodeficiency virus type 1 subtype C protease cleavage sites: an indication of viral fitness? J Virol 2003,77:9422-9430.

Department of Health & Social Security: Acquired Immune Deficiency Syndrome, general information for doctors. 1985.

de Roda Husman AM, van Rij RP, Blaak H, Broersen S, Schuitemaker H. Adaptation to promiscuous usage of chemokine receptors is not a prerequisite for human immunodeficiency virus type 1 disease progression. J Infect Dis 1999,180:1106-1115.

de Saizieu A, Gardes C, Flint N, et al. Microarray-based identification of a novel Streptococcus pneumoniae regulon controlled by an autoinduced peptide. J Bacteriol 2000,182:4696-4703.

De Wolf F, Hogervorst E, Goudsmit J, et al. Syncytium-inducing and non-syncytium-inducing capacity of human immunodeficiency virus type 1 subtypes other than B: phenotypic and genotypic characteristics. WHO Network for HIV Isolation and Characterization. AIDS Res Hum Retroviruses 1994,10:1387-1400.

Dean M, Carrington M, Winkler C, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study, Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco City Cohort, ALIVE Study. Science 1996,273:1856-1862.

DeBiasi RL, Clarke P, Meintzer S, et al. Reovirus-induced alteration in expression of apoptosis and DNA repair genes with potential roles in viral pathogenesis. J Virol 2003,77:8934-8947.

del Amo J, Petruckevitch A, Phillips A, et al. Disease progression and survival in HIV-1-infected Africans in London. Aids 1998,12:1203-1209.

del Real G, Jimenez-Baranda S, Mira E, et al. Statins inhibit HIV-1 infection by down-regulating Rho activity. J Exp Med 2004,200:541-547.

Delgado E, Thomson MM, Villahermosa ML, et al. Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. J Acquir Immune Defic Syndr 2002,29:536-543.

Delwart E, Magierowska M, Royz M, et al. Homogeneous quasispecies in 16 out of 17 individuals during very early HIV-1 primary infection. Aids 2002,16:189-195.

Delwart EL, Shpaer EG, Louwagie J, et al. Genetic relationships determined by a DNA heteroduplex mobility assay: analysis of HIV-1 env genes. Science 1993,262:1257-1261.

Deng H, Liu R, Ellmeier W, et al. Identification of a major co-receptor for primary isolates of HIV-1. Nature 1996,381:661-666.

Depienne C, Mousnier A, Leh H, et al. Characterization of the nuclear import pathway for HIV-1 integrase. J Biol Chem 2001,276:18102-18107.

Descamps D, Apetrei C, Collin G, Damond F, Simon F, Brun-Vezinet F. Naturally occurring decreased susceptibility of HIV-1 subtype G to protease inhibitors. Aids 1998,12:1109-1111.

Descamps D, Collin G, Letourneur F, et al. Susceptibility of human immunodeficiency virus type 1 group O isolates to antiretroviral agents: in vitro phenotypic and genotypic analyses. J Virol 1997,71:8893-8898.

Deubel V, Nogueira RM, Drouet MT, Zeller H, Reynes JM, Ha DQ. Direct sequencing of genomic cDNA fragments amplified by the polymerase chain reaction for molecular epidemiology of dengue-2 viruses. Arch Virol 1993,129:197-210.

Dittmar MT, Simmons G, Hibbitts S, et al. Langerhans cell tropism of human immunodeficiency virus type 1 subtype A through F isolates derived from different transmission groups. J Virol 1997,71:8008-8013.

Dolan KT, Twist EM, Horton-Slight P, et al. Epidemiology of rotavirus electropherotypes determined by a simplified diagnostic technique with RNA analysis. J Clin Microbiol 1985,21:753-758.

Domachowske JB, Bonville CA, Easton AJ, Rosenberg HF. Differential expression of proinflammatory cytokine genes in vivo in response to pathogenic and nonpathogenic pneumovirus infections. J Infect Dis 2002,186:8-14.

Doms RW, Trono D. The plasma membrane as a combat zone in the HIV battlefield. Genes Dev 2000,14:2677-2688.

Donaghy H, Gazzard B, Gotch F, Patterson S. Dysfunction and infection of freshly isolated blood myeloid and plasmacytoid dendritic cells in patients infected with HIV-1. Blood 2003,101:4505-4511.

Doranz BJ, Rucker J, Yi Y, et al. A dual-tropic primary HIV-1 isolate that uses fusin and the beta-chemokine receptors CKR-5, CKR-3, and CKR-2b as fusion cofactors. Cell 1996,85:1149-1158.

Douek DC, Picker LJ, Koup RA. T cell dynamics in HIV-1 infection. Annu Rev Immunol 2003,21:265-304.

Dragic T, Litwin V, Allaway GP, et al. HIV-1 entry into CD4+ cells is mediated by the chemokine receptor CC-CKR-5. Nature 1996,381:667-673.

du Bois RM, Branthwaite MA, Mikhail JR, Batten JC. Primary Pneumocystis carinii and cytomegalovirus infections. Lancet 1981,2:1339.

DuBridge RB, Tang P, Hsia HC, Leong PM, Miller JH, Calos MP. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. Mol Cell Biol 1987,7:379-387.

Duechler M, Skern T, Sommergruber W, et al. Evolutionary relationships within the human rhinovirus genus: comparison of serotypes 89, 2, and 14. Proc Natl Acad Sci U S A 1987,84:2605-2609.

Dworakowska B, Dolowy K. Ion channels-related diseases. Acta Biochim Pol 2000,47:685-703.

Eckmann L, Smith JR, Housley MP, Dwinell MB, Kagnoff MF. Analysis by high density cDNA arrays of altered gene expression in human intestinal epithelial cells in response to infection with the invasive enteric bacteria Salmonella. J Biol Chem 2000,275:14084-14094.

Edwards MC, Liegeois N, Horecka J, et al. Human CPR (cell cycle progression restoration) genes impart a Far- phenotype on yeast cells. Genetics 1997,147:1063-1076.

Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998,95:14863-14868.

Etienne-Manneville S, Hall A. Rho GTPases in cell biology. Nature 2002,420:629-635.

European Collaborative Study. Children born to women with HIV-1 infection: natural history and risk of transmission. Lancet. 1991,337:253-260.

Falk K, Gratama JW, Rowe M, et al. The role of repetitive DNA sequences in the size variation of Epstein-Barr virus (EBV) nuclear antigens, and the identification of different EBV isolates using RFLP and PCR analysis. J Gen Virol 1995,76 (Pt 4):779-790.

Feingold KR, Hardardottir I, Grunfeld C. Beneficial effects of cytokine induced hyperlipidemia. Z Ernahrungswiss 1998,37 Suppl 1:66-74.

Felsenstein J. PHYLIP—Phylogeny Inference Package. In. Version 3.2 ed: Cladistics 5:164–166; 1989.

Feng Y, Broder CC, Kennedy PE, Berger EA. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. Science 1996,272:872-877.

Fenyo EM, Morfeldt-Manson L, Chiodi F, et al. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. J Virol 1988,62:4414-4419.

Finnegan CM, Rawat SS, Puri A, Wang JM, Ruscetti FW, Blumenthal R. Ceramide, a target for antiretroviral therapy. Proc Natl Acad Sci U S A 2004,101:15452-15457.

Fletcher TM, 3rd, Brichacek B, Sharova N, et al. Nuclear import and cell cycle arrest functions of the HIV-1 Vpr protein are encoded by two separate genes in HIV-2/SIV(SM). Embo J 1996,15:6155-6165.

Fleury H, Recordon-Pinson P, Caumont A, et al. HIV type 1 diversity in France, 1999-2001: molecular characterization of non-B HIV type 1 subtypes and potential impact on susceptibility to antiretroviral drugs. AIDS Res Hum Retroviruses 2003,19:41-47.

Franke EK, Yuan HE, Luban J. Specific incorporation of cyclophilin A into HIV-1 virions. Nature 1994,372:359-362.

Frater AJ, Beardall A, Ariyoshi K, et al. Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. Aids 2001,15:1493-1502.

Freed EO, Ross SR. Retroviruses 2004: Review of the 2004 Cold Spring Harbor Retroviruses conference. Retrovirology 2004,1:25.

Fujioka S, Koide H, Kitaura Y, Duguchi H, Kawamura K. Analysis of enterovirus genotypes using single-strand conformation polymorphisms of polymerase chain reaction products. J Virol Methods 1995,51:253-258.

Gaggero A, Avendano LF, Fernandez J, Spencer E. Nosocomial transmission of rotavirus from patients admitted with diarrhea. J Clin Microbiol 1992,30:3294-3297.

Galai N, Kalinkovich A, Burstein R, Vlahov D, Bentwich Z. African HIV-1 subtype C and rate of progression among Ethiopian immigrants in Israel. Lancet 1997,349:180-181.

Gale CV, Myers R, Tedder RS, Williams IG, Kellam P. Development of a novel human immunodeficiency virus type 1 subtyping tool, Subtype Analyzer (STAR): analysis of subtype distribution in London. AIDS Res Hum Retroviruses 2004,20:457-464.

Gallay P, Hope T, Chin D, Trono D. HIV-1 infection of nondividing cells through the recognition of integrase by the importin/karyopherin pathway. Proc Natl Acad Sci U S A 1997,94:9825-9830.

Gallo RC, Poiesz BJ, Ruscetti FW. Regulation of human T-cell proliferation: T-cell growth factor and isolation of a new class of type-C retroviruses from human T-cells. Haematol Blood Transfus 1981,26:502-514.

Gallo RC, Salahuddin SZ, Popovic M, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science 1984,224:500-503.

Gao F, Robertson DL, Carruthers CD, et al. A comprehensive panel of near-full-length clones and reference sequences for non-subtype B isolates of human immunodeficiency virus type 1. J Virol 1998,72:5680-5698.

Gao F, Vidal N, Li Y, et al. Evidence of two distinct subsubtypes within the HIV-1 subtype A radiation. AIDS Res Hum Retroviruses 2001,17:675-688.

Gao F, Yue L, White AT, et al. Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. Nature 1992,358:495-499.

Gao Y, Paxinos E, Galovich J, et al. Characterization of a Subtype D Human Immunodeficiency Virus Type 1 Isolate That Was Obtained from an Untreated Individual and That Is Highly Resistant to Nonnucleoside Reverse Transcriptase Inhibitors. J. Virol. 2004,78:5390-5401.

Garber ME, Jones KA. HIV-1 Tat: coping with negative elongation factors. Curr Opin Immunol 1999,11:460-465.

Garcia-Martinez LF, Mavankal G, Neveu JM, Lane WS, Ivanov D, Gaynor RB. Purification of a Tat-associated kinase reveals a TFIIH complex that modulates HIV-1 transcription. Embo J 1997,16:2836-2850.

Garrus JE, von Schwedler UK, Pornillos OW, et al. Tsg101 and the vacuolar protein sorting pathway are essential for HIV-1 budding. Cell 2001,107:55-65.

Geijtenbeek TB, Torensma R, van Vliet SJ, et al. Identification of DC-SIGN, a novel dendritic cell-specific ICAM-3 receptor that supports primary immune responses. Cell 2000,100:575-585.

Geiss G, Jin G, Guo J, Bumgarner R, Katze MG, Sen GC. A comprehensive view of regulation of gene expression by double-stranded RNA-mediated cell signaling. J Biol Chem 2001,276:30178-30182.

Geiss GK, Bumgarner RE, An MC, et al. Large-scale monitoring of host cell gene expression during HIV-1 infection using cDNA microarrays. Virology 2000,266:8-16.

Gervaix A, West D, Leoni LM, Richman DD, Wong-Staal F, Corbeil J. A new reporter cell line to monitor HIV infection and drug susceptibility in vitro. Proc Natl Acad Sci U S A 1997,94:4653-4658.

Geyer M, Yu H, Mandic R, et al. Subunit H of the V-ATPase binds to the medium chain of adaptor protein complex 2 and connects Nef to the endocytic machinery. J Biol Chem 2002,277:28521-28529.

Ghosh P, Dahms NM, Kornfeld S. Mannose 6-phosphate receptors: new twists in the tale. Nat Rev Mol Cell Biol 2003,4:202-212.

Gibellini D, Re MC, La Placa M, Zauli G. Differentially expressed genes in HIV-1 tat-expressing CD4(+) T-cell line. Virus Res 2002,90:337-345.

Gibellini D, Re MC, Ponti C, et al. Extracellular Tat activates c-fos promoter in low serum-starved CD4+ T cells. Br J Haematol 2001,112:663-670.

Giffin MJ, Stroud JC, Bates DL, von Koenig KD, Hardin J, Chen L. Structure of NFAT1 bound as a dimer to the HIV-1 LTR kappa B element. Nat Struct Biol 2003,10:800-806.

Giguere JF, Tremblay MJ. Statin compounds reduce human immunodeficiency virus type 1 replication by preventing the interaction between virion-associated host intercellular adhesion molecule 1 and its natural cell surface ligand LFA-1. J Virol 2004,78:12062-12065.

Giraud S, Diaz-Latoud C, Hacot S, Textoris J, Bourette RP, Diaz JJ. US11 of herpes simplex virus type 1 interacts with HIPK2 and antagonizes HIPK2-induced cell growth arrest. J Virol 2004,78:2984-2993.

Gnaore E, De Cock KM, Gayle H, et al. Prevalence of and mortality from HIV type 2 in Guinea Bissau, West Africa. Lancet 1989,2:513.

Goedert JJ, Gallo RC. Epidemiological evidence that HTLV-III is the AIDS agent. Eur J Epidemiol 1985,1:155-159.

Goedert JJ, Neuland CY, Wallen WC, et al. Amyl nitrite may alter T lymphocytes in homosexual men. Lancet 1982,1:412-416.

Goff SP. Intracellular trafficking of retroviral genomes during the early phase of infection: viral exploitation of cellular pathways. J Gene Med 2001,3:517-528.

Gomes D, Goncalves MF. Different pathways to nelfinavir genotypic resistance in HIV-1 subtypes B and G. 9th Conference on Retroviruses and Opportunistic Infections. Seattle, WA, 2428 2002.

Gomes LI, Silva RL, Stolf BS, et al. Comparative analysis of amplified and nonamplified RNA for hybridization in cDNA microarray. Anal Biochem 2003,321:244-251.

Gonzalez E, Kulkarni H, Bolivar H, et al. The Influence of CCL3L1 Gene-Containing Segmental Duplications on HIV-1/AIDS Susceptibility. Science 2005.

Gonzalez-Villasenor LI, Wu K, Yang L. A solid phase plate assay for HIV-1 genotyping. Mol Cell Probes 2000,14:137-147.

Goodbourn S, Didcock L, Randall RE. Interferons: cell signalling, immune modulation, antiviral response and virus countermeasures. J Gen Virol 2000,81:2341-2364.

Goodrich DW, Duesberg PH. Retroviral recombination during reverse transcription. Proc Natl Acad Sci U S A 1990,87:2052-2056.

Gordeuk VR, Delanghe JR, Langlois MR, Boelaert JR. Iron status and the outcome of HIV infection: an overview. J Clin Virol 2001,20:111-115.

Gottlieb MS, Schroff R, Schanker HM, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. N Engl J Med 1981,305:1425-1431.

Gottlinger HG. The HIV-1 assembly machine. Aids 2001,15 Suppl 5:S13-20.

Goulder PJ, Phillips RE, Colbert RA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. Nat Med 1997,3:212-217.

Greenberg ME, Bronson S, Lock M, Neumann M, Pavlakis GN, Skowronski J. Co-localization of HIV-1 Nef with the AP-2 adaptor protein complex correlates with Nef-induced CD4 down-regulation. Embo J 1997,16:6964-6976.

Griffin SV, Hiromura K, Pippin J, et al. Cyclin-Dependent Kinase 5 Is a Regulator of Podocyte Differentiation, Proliferation, and Morphology. Am J Pathol 2004,165:1175-1185.

Grossman Z, Paul WE. The impact of HIV on naive T-cell homeostasis. Nat Med 2000,6:976-977.

Grossman Z, Vardinon N, Chemtob D, et al. Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B. Aids 2001,15:1453-1460.

Guadalupe M, Reay E, Sankaran S, et al. Severe CD4+ T-cell depletion in gut lymphoid tissue during primary human immunodeficiency virus type 1 infection and substantial delay in restoration following highly active antiretroviral therapy. J Virol 2003,77:11708-11717.

Guerra S, Lopez-Fernandez LA, Conde R, Pascual-Montano A, Harshman K, Esteban M. Microarray analysis reveals characteristic changes of host cell gene expression in response to attenuated modified vaccinia virus Ankara infection of human HeLa cells. J Virol 2004,78:5820-5834.

Gunson HH. Trends in blood transfusion practice in England and Wales. Health Trends 1986,18:76-79.

Gupta P, Collins KB, Ratner D, et al. Memory CD4(+) T cells are the earliest detectable human immunodeficiency virus type 1 (HIV-1)-infected cells in the female genital mucosal tissue during HIV-1 transmission in an organ culture system. J Virol 2002,76:9868-9876.

Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids. Symp. Ser. 1999,41:95-98.

Halsey NA, Boulos R, Holt E, et al. Transmission of HIV-1 infections from mothers to infants in Haiti. Impact on childhood mortality and malnutrition. The CDS/JHU AIDS Project Team. Jama 1990,264:2088-2092.

Hanley TM, Kiefer HLB, Schnitzler AC, Marcello JE, Viglianti GA. Retinoid-Dependent Restriction of Human Immunodeficiency Virus Type 1 Replication in Monocytes/Macrophages. J. Virol. 2004,78:2819-2830.

Hansasuta P, Rowland-Jones SL. HIV-1 transmission and acute HIV-1 infection. Br Med Bull 2001,58:109-127.

Hardie DR, Kannemeyer J, Stannard LM. DNA single strand conformation polymorphism identifies five defined strains of hepatitis B virus (HBV) during an outbreak of HBV infection in an oncology unit. J Med Virol 1996,49:49-54.

Harris RS, Bishop KN, Sheehy AM, et al. DNA deamination mediates innate immunity to retroviral infection. Cell 2003,113:803-809.

Harris RS, and Liddament MT. Retroviral restriction by APOBEC proteins. Nat Rev Immunol. 2004 Nov;4 (11):868-77.

Hartigan JA. Clustering. Annu Rev Biophys Bioeng 1973,2:81-101.

Hasegawa I, Tanaka Y, Kramvis A, et al. Novel hepatitis B virus genotype a subtyping assay that distinguishes subtype Aa from Ae and its application in epidemiological studies. J Virol 2004,78:7575-7581.

Hayward RE, Derisi JL, Alfadhli S, Kaslow DC, Brown PO, Rathod PK. Shotgun DNA microarrays and stage-specific gene expression in Plasmodium falciparum malaria. Mol Microbiol 2000,35:6-14.

He J, Choe S, Walker R, Di Marzio P, Morgan DO, Landau NR. Human immunodeficiency virus type 1 viral protein R (Vpr) arrests cells in the G2 phase of the cell cycle by inhibiting p34cdc2 activity. J Virol 1995,69:6705-6711.

Hierholzer JC. Adenoviruses in the immunocompromised host. Clin Microbiol Rev 1992,5:262-274.

Hierholzer JC, Wigand R, Anderson LJ, Adrian T, Gold JW. Adenoviruses from patients with AIDS: a plethora of serotypes and a description of five new serotypes of subgenus D (types 43-47). J Infect Dis 1988,158:804-813.

Hill JM, Lukiw WJ, Gebhardt BM, et al. Gene expression analyzed by microarrays in HSV-1 latent mouse trigeminal ganglion following heat stress. Virus Genes 2001,23:273-280.

Hindmarsh P, Leis J. Retroviral DNA integration. Microbiol Mol Biol Rev 1999,63:836-843

Hirsch MS, Richman DD. The role of genotypic resistance testing in selecting therapy for HIV. Jama 2000,284:1649-1650.

Hobbs WE, 2nd, DeLuca NA. Perturbation of cell cycle progression and cellular gene expression as a function of herpes simplex virus ICP0. J Virol 1999,73:8245-8255.

Holmes I, Durbin R. Dynamic programming alignment accuracy. J Comput Biol 1998,5:493-504.

Holmes RDMPEC. Molecular evolution : a phylogenetic approach: Oxford ; Malden, MA : Blackwell Science; 1998.

Honda M, Kaneko S, Kawai H, Shirota Y, Kobayashi K. Differential gene expression between chronic hepatitis B and C hepatic lesion. Gastroenterology 2001,120:955-966.

Hooker CW, Scott J, Apolloni A, Parry E, Harrich D. Human immunodeficiency virus type 1 reverse transcription is stimulated by tat from other lentiviruses. Virology 2002,300:226-235.

Horton JD. Sterol regulatory element-binding proteins: transcriptional activators of lipid synthesis. Biochem Soc Trans 2002,30:1091-1095.

Howard TM, Rasheed S. Genomic structure and nucleotide sequence analysis of a new HIV type 1 subtype A strain from Nigeria. AIDS Res Hum Retroviruses 1996,12:1413-1425.

Hsia SC, Shi YB. Chromatin disruption and histone acetylation in regulation of the human immunodeficiency virus type 1 long terminal repeat by thyroid hormone receptor. Mol Cell Biol 2002,22:4043-4052.

Hu DJ, Buve A, Baggs J, van der Groen G, Dondero TJ. What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. Aids 1999,13:873-881.

Hu DJ, Vanichseni S, Mastro TD, et al. Viral load differences in early infection with two HIV-1 subtypes. Aids 2001,15:683-691.

Hu L, Wang J, Baggerly K, et al. Obtaining reliable information from minute amounts of RNA using cDNA microarrays. BMC Genomics 2002,3:16.

Huang ES, Alford CA, Reynolds DW, Stagno S, Pass RF. - a Molecular epidemiology of cytomegalovirus infections in women and their infants. N Engl J Med 1980,303:958-962.

Huang ES, Huong SM, Tegtmeier GE, Alford C. - b Cytomegalovirus: genetic variation of viral genomes. Ann N Y Acad Sci 1980,354:332-346.

Huang X, Seifert U, Salzmann U, et al. The RTP site shared by the HIV-1 Tat protein and the 11S regulator subunit alpha is crucial for their effects on proteasome function including antigen processing. J Mol Biol 2002,323:771-782.

Huang Y, Paxton WA, Wolinsky SM, et al. The role of a mutant CCR5 allele in HIV-1 transmission and disease progression. Nat Med 1996,2:1240-1243.

Hue S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. Proc Natl Acad Sci U S A. 2005 Mar 22;102(12):4425-9.

Hughes PJ, North C, Jellis CH, Minor PD, Stanway G. The nucleotide sequence of human rhinovirus 1B: molecular relationships within the rhinovirus genus. J Gen Virol 1988,69 (Pt 1):49-58.

Hunziker W, Geuze HJ. Intracellular trafficking of lysosomal membrane proteins. Bioessays 1996,18:379-389.

IHGSC. Finishing the euchromatic sequence of the human genome. Nature 2004,431:931-945.

Iizuka N, Oka M, Yamada-Okabe H, et al. Comparison of gene expression profiles between hepatitis B virus- and hepatitis C virus-infected hepatocellular carcinoma by oligonucleotide microarray data on the basis of a supervised learning method. Cancer Res 2002,62:3939-3944.

Ingley E, Chappell D, Poon SY, et al. Thyroid hormone receptor-interacting protein 1 modulates cytokine and nuclear hormone signaling in erythroid cells. J Biol Chem 2001,276:43428-43434.

Ishizuka T, Satoh T, Monden T, et al. Human immunodeficiency virus type 1 Tat binding protein-1 is a transcriptional coactivator specific for TR. Mol Endocrinol 2001,15:1329-1343.

Iyer VR, Eisen MB, Ross DT, et al. The transcriptional program in the response of human fibroblasts to serum. Science 1999,283:83-87.

Izmailova E, Bertley FM, Huang Q, et al. HIV-1 Tat reprograms immature dendritic cells to express chemoattractants for activated T cells and macrophages. Nat Med 2003,9:191-197.

Janket ML, Manickam P, Majumder B, et al. Differential regulation of host cellular genes by HIV-1 viral protein R (Vpr): cDNA microarray analysis using isogenic virus. Biochem Biophys Res Commun 2004,314:1126-1132.

Jazwinski SM, Conzelmann A. LAG1 puts the focus on ceramide signaling. Int J Biochem Cell Biol 2002,34:1491-1495.

Jeeninga RE, Hoogenkamp M, Armand-Ugon M, de Baar M, Verhoef K, Berkhout B. Functional differences between the long terminal repeat transcriptional promoters of human immunodeficiency virus type 1 subtypes A through G. J Virol 2000,74:3740-3751.

Jensen MA, Li FS, van 't Wout AB, et al. Improved coreceptor usage prediction and genotypic monitoring of R5-to-X4 transition by motif analysis of human immunodeficiency virus type 1 env V3 loop sequences. J Virol 2003,77:13376-13388.

Jensen MA, van 't Wout AB. Predicting HIV-1 coreceptor usage with sequence analysis. AIDS Rev 2003,5:104-112.

Jenson SD, Robetorye RS, Bohling SD, et al. Validation of cDNA microarray gene expression data obtained from linearly amplified RNA. Mol Pathol 2003,56:307-312.

Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 2000,74:1234-1240.

Jiang C, Ito M, Piening V, Bruck K, Roeder RG, Xiao H. TIP30 interacts with an estrogen receptor alpha-interacting coactivator CIA and regulates c-myc transcription. J Biol Chem 2004,279:27781-27789.

Johannes G, Carter MS, Eisen MB, Brown PO, Sarnow P. Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. Proc Natl Acad Sci U S A 1999,96:13118-13123.

Jowett JB, Planelles V, Poon B, Shah NP, Chen ML, Chen IS. The human immunodeficiency virus type 1 vpr gene arrests infected T cells in the G2 + M phase of the cell cycle. J Virol 1995,69:6304-6313.

Kaleebu P, French N, Mahe C, et al. Effect of human immunodeficiency virus (HIV) type 1 envelope subtypes A and D on disease progression in a large cohort of HIV-1-positive persons in Uganda. J Infect Dis 2002,185:1244-1250.

Kaleebu P, Ross A, Morgan D, et al. Relationship between HIV-1 Env subtypes A and D and disease progression in a rural Ugandan cohort. Aids 2001,15:293-299.

Kalyanaraman VS, Sarngadharan MG, Robert-Guroff M, Miyoshi I, Golde D, Gallo RC. A new subtype of human T-cell leukemia virus (HTLV-II) associated with a T-cell variant of hairy cell leukemia. Science 1982,218:571-573.

Kamradt T, Niese D, Vogel F. Slim disease (AIDS). Lancet 1985,2:1425.
Kanki PJ, Hamel DJ, Sankale JL, et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. J Infect Dis 1999,179:68-73.

Kantor R, Zijenah LS, Shafer RW, et al. HIV-1 subtype C reverse transcriptase and protease genotypes in Zimbabwean patients failing antiretroviral therapy. AIDS Res Hum Retroviruses 2002,18:1407-1413.

Kaplan AH. Assembly of the HIV-1 core particle. AIDS Rev 2002,4:104-111.

Kari C, Chan TO, Rocha de Quadros M, Rodeck U. Targeting the epidermal growth factor receptor in cancer: apoptosis takes center stage. Cancer Res 2003,63:1-5.

Karlsson A, Parsmyr K, Sandstrom E, Fenyo EM, Albert J. MT-2 cell tropism as prognostic marker for disease progression in human immunodeficiency virus type 1 infection. J Clin Microbiol 1994,32:364-370.

Kartvelishvili A, Lesner A, Szponar M, Simm M. Microarray analysis of differentially expressed genes in cells resistant to HIV-1. Immunol Lett 2004,93:79-86.

Kaslow RA, Dorak T, Tang JJ. Influence of host genetic variation on susceptibility to HIV type 1 infection. J Infect Dis 2005,191 Suppl 1:S68-77.

Katayama Y, Shibahara K, Kohama T, Homma M, Hotta H. Molecular epidemiology and changing distribution of genotypes of measles virus field strains in Japan. J Clin Microbiol 1997,35:2651-2653.

Kato-Maeda M, Rhee JT, Gingeras TR, et al. Comparing genomes within the species Mycobacterium tuberculosis. Genome Res 2001,11:547-554.

Katz BZ, Niederman JC, Olson BA, Miller G. Fragment length polymorphisms among independent isolates of Epstein-Barr virus from immunocompromised and normal hosts. J Infect Dis 1988,157:299-308.

Katzenstein TL, Pedersen C, Nielsen C, Lundgren JD, Jakobsen PH, Gerstoft J. Longitudinal serum HIV RNA quantification: correlation to viral phenotype at seroconversion and clinical outcome. Aids 1996,10:167-173.

Kawamura T, Cohen SS, Borris DL, et al. Candidate microbicides block HIV-1 infection of human immature Langerhans cells within epithelial tissue explants. J Exp Med 2000,192:1491-1500.

Kedzierska K, Ellery P, Mak J, Lewin SR, Crowe SM, Jaworowski A. HIV-1 down-modulates gamma signaling chain of Fc gamma R in human macrophages: a possible mechanism for inhibition of phagocytosis. J Immunol 2002,168:2895-2903.

Keet IP, Janssen M, Veugelers PJ, et al. Longitudinal analysis of CD4 T cell counts, T cell reactivity, and human immunodeficiency virus type 1 RNA levels in persons remaining AIDS-free despite CD4 cell counts <200 for >5 years. J Infect Dis 1997,176:665-671.

Kenyon G. Resistance study to re-evaluate HAART. Nat Med 2001,7:515.

Kenzelmann M, Klaren R, Hergenhahn M, et al. High-accuracy amplification of nanogram total RNA amounts for gene profiling. Genomics 2004,83:550-558.

Kerkau T, Bacik I, Bennink JR, et al. The human immunodeficiency virus type 1 (HIV-1) Vpu protein interferes with an early step in the biosynthesis of major histocompatibility complex (MHC) class I molecules. J Exp Med 1997,185:1295-1305.

Kerr JR, Curran MD, Moore JE, et al. Genetic diversity in the non-structural gene of parvovirus B19 detected by single-stranded conformational polymorphism assay (SSCP) and partial nucleotide sequencing. J Virol Methods 1995,53:213-222.

Kerr MK, Martin M, Churchill GA. Analysis of variance for gene expression microarray data. J Comput Biol 2000,7:819-837.

Kessler N, Ferraris O, Palmer K, Marsh W, Steel A. Use of the DNA flow-thru chip, a three-dimensional biochip, for typing and subtyping of influenza viruses. J Clin Microbiol 2004,42:2173-2185.

Khaiboullina SF, Rizvanov AA, Otteson E, Miyazato A, Maciejewski J, St Jeor S. Regulation of cellular gene expression in endothelial cells by sin nombre and prospect hill viruses. Viral Immunol 2004,17:234-251.

Khodarev NN, Advani SJ, Gupta N, Roizman B, Weichselbaum RR. Accumulation of specific RNAs encoding transcriptional factors and stress response proteins against a background of severe depletion of cellular RNAs in cells infected with herpes simplex virus 1. Proc Natl Acad Sci U S A 1999,96:12062-12067.

Kiepiela P, Leslie AJ, Honeyborne I, et al. Dominant influence of HLA-B in mediating the potential co-evolution of HIV and HLA. Nature 2004,432:769-775.

Kilpatrick BA, Huang ES, Pagano JS. Analysis of cytomegalovirus genomes with restriction endonucleases Hin D III and EcoR-1. J Virol 1976,18:1095-1105.

Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 1980,16:111-120.

Kingman Ca. The search for the virus, the scientific discovery of AIDS and the quest for a cure: Penguin Books; 1988.

Kino T, Pavlakis GN. Partner molecules of accessory protein Vpr of the human immunodeficiency virus type 1. DNA Cell Biol 2004,23:193-205.

Klatzmann D, Champagne E, Chamaret S, et al. T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. Nature 1984,312:767-768.

Kobayashi M, Takaori-Kondo A, Shindo K, Abudu A, Fukunaga K, Uchiyama T. APOBEC3G targets specific virus species. J Virol 2004,78:8238-8244.

Koch G, Benz C, Schmidt G, Olenik C, Aktories K. Role of Rho protein in lovastatin-induced breakdown of actin cytoskeleton. J Pharmacol Exp Ther 1997,283:901-909.

Komada H, Klippmark E, Orvell C, Randall RE, Ito Y, Norrby E. Immunological relationships between parainfluenza virus type 4 and other paramyxoviruses studied by use of monoclonal antibodies. Arch Virol 1991,116:277-283.

Koonin EV. Genome sequences: genome sequence of a model prokaryote. Curr Biol 1997,7:R656-659.

Koot M, Keet IP, Vos AH, et al. Prognostic value of HIV-1 syncytium-inducing phenotype for rate of CD4+ cell depletion and progression to AIDS. Ann Intern Med 1993,118:681-688.

Korber B, Muldoon M, Theiler J, et al. Timing the ancestor of the HIV-1 pandemic strains. Science 2000,288:1789-1796.

Koulinska IN, Ndung'u T, Mwakagile D, et al. A new human immunodeficiency virus type 1 circulating recombinant form from Tanzania. AIDS Res Hum Retroviruses 2001,17:423-431.

Kuan MM, Wang GR. Identification of a measles virus isolate from a recent outbreak in northern Taiwan. Proc Natl Sci Counc Repub China B 1997,21:96-99.

Kumar A, Manna SK, Dhawan S, Aggarwal BB. HIV-Tat Protein Activates c-Jun N-Terminal Kinase and Activator Protein-1. J Immunol 1998,161:776-781.

Kumar S, Subramanian S. Mutation rates in mammalian genomes. Proc Natl Acad Sci U S A 2002,99:803-808.

Kunanusont C, Foy HM, Kreiss JK, et al. HIV-1 subtypes and male-to-female transmission in Thailand. Lancet 1995,345:1078-1083.

Kurosu T, Mukai T, Komoto S, et al. Human immunodeficiency virus type 1 subtype C exhibits higher transactivation activity of Tat than subtypes B and E. Microbiol Immunol 2002,46:787-799.

Lafrenie RM, Wahl LM, Epstein JS, Hewlett IK, Yamada KM, Dhawan S. HIV-1-Tat modulates the function of monocytes and alters their interactions with microvessel endothelial cells. A mechanism of HIV pathogenesis. J Immunol 1996,156:1638-1645.

Lareu RR, Swanson NR, Fox SA. Rapid and sensitive genotyping of hepatitis C virus by single-strand conformation polymorphism. J Virol Methods 1997,64:11-18.

Laukkanen T, Carr JK, Janssens W, et al. Virtually full-length subtype F and F/D recombinant HIV-1 from Africa and South America. Virology 2000,269:95-104.

Le Pogam S, Dubois F, Christen R, Raby C, Cavicchini A, Goudeau A. Comparison of DNA enzyme immunoassay and line probe assays (Inno-LiPA HCV I and II) for hepatitis C virus genotyping. J Clin Microbiol 1998,36:1461-1463.

Lecossier D, Bouchonnet F, Clavel F, Hance AJ. Hypermutation of HIV-1 DNA in the absence of the Vif protein. Science 2003,300:1112.

Lee ML, Kuo FC, Whitmore GA, Sklar J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc Natl Acad Sci U S A 2000,97:9834-9839.

Lee WM. Hepatitis B virus infection. N Engl J Med 1997,337:1733-1745.

Leitner T, Foley, B, Hahn, B, Marx, P, McCutchan, F, Mellors, J, Wolinsky, S, and Korber, B. HIV Sequence Compendium: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, LA-UR number 04-7420; 2001.

Lennon GG, Lehrach H. Hybridization analyses of arrayed cDNA libraries. Trends Genet 1991,7:314-317.

Leroux C, Le Provost F, Petit E, Bernard L, Chilliard Y, Martin P. Real-time RT-PCR and cDNA macroarray to study the impact of the genetic polymorphism at the alphas1-casein locus on the expression of genes in the goat mammary gland during lactation. Reprod Nutr Dev 2003,43:459-469.

Leslie AJ, Pfafferott KJ, Chetty P, et al. HIV evolution: CTL escape mutation and reversion after transmission. Nat Med 2004,10:282-289.

Leung YF, Cavalieri D. Fundamentals of cDNA microarray data analysis. Trends Genet 2003,19:649-659.

Levett PN, Gu M, Luan B, et al. Longitudinal study of molecular epidemiology of small round-structured viruses in a pediatric population. J Clin Microbiol 1996,34:1497-1501.

Levy JA, Hoffman AD, Kramer SM, Landis JA, Shimabukuro JM, Oshiro LS. Isolation of lymphocytopathic retroviruses from San Francisco patients with AIDS. Science 1984,225:840-842.

Levy R, Jais JP, Tourani JM, Even P, Andrieu JM. Long-term follow-up of HIV positive asymptomatic patients having received cyclosporin A. Adv Exp Med Biol 1995,374:229-234.

Li J, Adams L, Schwartz SM, Bumgamer RE. RNA amplification, fidelity and reproducibility of expression profiling. C R Biol 2003,326:1021-1030.

Li J, Chen S, Evans DH. Typing and subtyping influenza virus using DNA microarrays and multiplex reverse transcriptase PCR. J Clin Microbiol 2001,39:696-704.

Li P, Stephenson AJ, Kuiper LJ, Burrell CJ. Double-stranded strong-stop DNA and the second template switch in human immunodeficiency virus (HIV) DNA synthesis. Virology 1993,194:82-88.

Li Q, Schacker T, Carlis J, Beilman G, Nguyen P, Haase AT. - a Functional genomic analysis of the response of HIV-1-infected lymphatic tissue to antiretroviral therapy. J Infect Dis 2004,189:572-582.

Li QG, Zheng QJ, Liu YH, Wadell G. Molecular epidemiology of adenovirus types 3 and 7 isolated from children with pneumonia in Beijing. J Med Virol 1996,49:170-177.

Li W, Simarro M, Kedersha N, Anderson P. – c FAST is a survival protein that senses mitochondrial stress and modulates TIA-1-regulated changes in protein expression. Mol Cell Biol 2004,24:10718-10732.

Li WH, Tanimura M, Sharp PM. Rates and dates of divergence between AIDS virus nucleotide sequences. Mol Biol Evol 1988,5:313-330.

Li Y, Li T, Liu S, et al. - b Systematic comparison of the fidelity of aRNA, mRNA and T-RNA on gene expression profiling using cDNA microarray. J Biotechnol 2004,107:19-28.

Liao Z, Cimakasky LM, Hampton R, Nguyen DH, Hildreth JE. Lipid rafts and HIV pathogenesis: host membrane cholesterol is required for infection by HIV type 1. AIDS Res Hum Retroviruses 2001,17:1009-1019.

Liitsola K, Tashkinova I, Laukkanen T, et al. HIV-1 genetic subtype A/B recombinant strain causing an explosive epidemic in injecting drug users in Kaliningrad. Aids 1998,12:1907-1919.

Little RF. AIDS-related non-Hodgkin's lymphoma: etiology, epidemiology, and impact of highly active antiretroviral therapy. Leuk Lymphoma 2003,44 Suppl 3:S63-68.

Liu LX, Heveker N, Fackler OT, et al. Mutation of a conserved residue (D123) required for oligomerization of human immunodeficiency virus type 1 Nef protein abolishes interaction with human thioesterase and results in impairment of Nef biological functions. J Virol 2000,74:5310-5319.

Liu R, Paxton WA, Choe S, et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. Cell 1996,86:367-377.

Lockhart DJ, Dong H, Byrne MC, et al. Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol 1996,14:1675-1680.

Lopez-Galindez C, Lopez JA, Melero JA, et al. Analysis of genetic variability and mapping of point mutations in influenza virus by the RNase A mismatch cleavage method. Proc Natl Acad Sci U S A 1988,85:3522-3526.

Lori F, Hall L, Lusso P, et al. Effect of reciprocal complementation of two defective human immunodeficiency virus type 1 (HIV-1) molecular clones on HIV-1 cell tropism and virulence. J Virol 1992,66:5553-5560.

Luciw PA, Potter SJ, Steimer K, Dina D, Levy JA. Molecular cloning of AIDS-associated retrovirus. Nature 1984,312:760-763.

Lukashov VV, Huismans R, Rakhmanova AG, et al. Circulation of subtype A and gagA/envB recombinant HIV type 1 strains among injecting drug users in St. Petersburg, Russia, correlates with geographical origin of infections. AIDS Res Hum Retroviruses 1999,15:1577-1583.

Lusic M, Marcello A, Cereseto A, Giacca M. Regulation of HIV-1 gene expression by histone acetylation and factor recruitment at the LTR promoter. Embo J 2003,22:6550-6561.

Machuca A, Ding L, Taffs R, et al. HIV type 2 primary isolates induce a lower degree of apoptosis "in vitro" compared with HIV type 1 primary isolates. AIDS Res Hum Retroviruses 2004,20:507-512.

Malim MH, Cullen BR. Rev and the fate of pre-mRNA in the nucleus: implications for the regulation of RNA processing in eukaryotes. Mol Cell Biol 1993,13:6180-6189.

Mangeat B, Turelli P, Caron G, Friedli M, Perrin L, Trono D. Broad antiretroviral defence by human APOBEC3G through lethal editing of nascent reverse transcripts. Nature 2003,424:99-103.

Mann DA, Mikaelian I, Zemmel RW, et al. A molecular rheostat. Co-operative rev binding to stem I of the rev-response element modulates human immunodeficiency virus type-1 late gene expression. J Mol Biol 1994,241:193-207.

Marathi UK, Dahlen M, Sunnerhagen P, et al. RAD1, a human structural homolog of the Schizosaccharomyces pombe RAD1 cell cycle checkpoint gene. Genomics 1998,54:344-347.

Mariani R, Chen D, Schrofelbauer B, et al. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. Cell 2003,114:21-31.

Marin M, Rose KM, Kozak SL, Kabat D. HIV-1 Vif protein binds the editing enzyme APOBEC3G and induces its degradation. Nat Med 2003,9:1398-1403.

Marlink R, Kanki P, Thior I, et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science 1994,265:1587-1590.

Marozsan AJ, Fraundorf E, Abraha A, et al. Relationships between infectious titer, capsid protein levels, and reverse transcriptase activities of diverse human immunodeficiency virus type 1 isolates. J Virol 2004,78:11130-11141.

Marx JL. New disease baffles medical community. Science 1982,217:618-621.

Marx JL. A virus by any other name. Science 1985,227:1449-1451.

Mastro TD, de Vincenzi I. Probabilities of sexual HIV-1 transmission. Aids 1996,10 Suppl A:S75-82.

Mastro TD, Kitayaporn D. HIV type 1 transmission probabilities: estimates from epidemiological studies. AIDS Res Hum Retroviruses 1998,14 Suppl 3:S223-227.

Mastro TD, Satten GA, Nopkesorn T, Sangkharomya S, Longini IM, Jr. Probability of female-to-male transmission of HIV-1 in Thailand. Lancet 1994,343:204-207.

Masur H, Michelis MA, Greene JB, et al. An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction. N Engl J Med 1981,305:1431-1438.

Matern HT, Yeaman C, Nelson WJ, Scheller RH. The Sec6/8 complex in mammalian cells: characterization of mammalian Sec3, subunit interactions, and expression of subunits in polarized cells. Proc Natl Acad Sci U S A 2001,98:9648-9653.

Matheson PB, Weedon J, Cappelli M, et al. Comparison of methods of estimating the mother-to-child transmission rate of human immunodeficiency virus type 1 (HIV-1). New York City Perinatal HIV Transmission Collaborative Study Group. Am J Epidemiol 1995,142:714-718.

Maury W. Regulation of equine infectious anemia virus expression. J Biomed Sci 1998,5:11-23.

Mayne M, Cheadle C, Soldan SS, et al. Gene expression profile of herpesvirus-infected T cells obtained using immunomicroarrays: induction of proinflammatory mechanisms. J Virol 2001,75:11641-11650.

Maziere JC, Landureau JC, Giral P, et al. Lovastatin inhibits HIV-1 expression in H9 human T lymphocytes cultured in cholesterol-poor medium. Biomed Pharmacother 1994,48:63-67.

McCormick-Davis C, Dalton SB, Singh DK, Stephens EB. Comparison of Vpu sequences from diverse geographical isolates of HIV type 1 identifies the presence of highly variable domains, additional invariant amino acids, and a signature sequence motif common to subtype C isolates. AIDS Res Hum Retroviruses 2000,16:1089-1095.

McCutchan FE. Understanding the genetic diversity of HIV-1. Aids 2000,14 Suppl 3:S31-44.

McNearney T, Hornickova Z, Markham R, et al. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc Natl Acad Sci U S A 1992,89:10247-10251.

Means RE, Matthews T, Hoxie JA, Malim MH, Kodama T, Desrosiers RC. Ability of the V3 Loop of Simian Immunodeficiency Virus To Serve as a Target for Antibody-Mediated Neutralization: Correlation of Neutralization Sensitivity, Growth in Macrophages, and Decreased Dependence on CD4. J. Virol. 2001,75:3903-3915.

Mellors JW, Munoz A, Giorgi JV, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Ann Intern Med 1997,126:946-954.

Meyer BE, Malim MH. The HIV-1 Rev trans-activator shuttles between the nucleus and the cytoplasm. Genes Dev 1994,8:1538-1547.

Mikovits J, Ruscetti F, Zhu W, Bagni R, Dorjsuren D, Shoemaker R. Potential cellular signatures of viral infections in human hematopoietic cells. Dis Markers 2001,17:173-178.

Miller ED, Smith JA, Lichtinger M, Wang L, Su L. Activation of the signal transducer and activator of transcription 1 signaling pathway in thymocytes from HIV-1-infected human thymus. Aids 2003,17:1269-1277.

Miller MD, Farnet CM, Bushman FD. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. J Virol 1997,71:5382-5390.
384. Miller V. European guidelines on resistance testing: why are they needed? HIV Med 2000,1:129-131.

Minor PD, Schild GC, Ferguson M, et al. Genetic and antigenic variation in type 3 polioviruses: characterization of strains by monoclonal antibodies and T1 oligonucleotide mapping. J Gen Virol 1982,61 (Pt 2):167-176.

Miyake H, Iizawa Y, Baba M. Novel reporter T-cell line highly susceptible to both CCR5- and CXCR4-using human immunodeficiency virus type 1 and its application to drug susceptibility tests. J Clin Microbiol 2003,41:2515-2521.

MMWR Weekly Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men-New York City and California. MMWR Morb Mortal Wkly Rep 1981,30:305-308.

MMWR Weekly-a. A cluster of Kaposi's sarcoma and Pneumocystis carinii pneumonia among homosexual male residents of Los Angeles and Orange Counties, California. MMWR Morb Mortal Wkly Rep 1982,31:305-307.

MMWR Weekly-b. Unexplained immunodeficiency and opportunistic infections in infants--New York, New Jersey, California. MMWR Morb Mortal Wkly Rep 1982,31:665-667.

MMWR Weekly Update: acquired immunodeficiency syndrome-Europe. MMWR Morb Mortal Wkly Rep 1985,34:147-150, 155-146.

Mocroft A, Lundgren JD. Starting highly active antiretroviral therapy: why, when and response to HAART. J Antimicrob Chemother 2004,54:10-13.

Mohammad-Panah R, Harrison R, Dhani S, et al. The Chloride Channel ClC-4 Contributes to Endosomal Acidification and Trafficking. J. Biol. Chem. 2003,278:29267-29277.

Montagnier L. Historical essay. A history of HIV discovery. Science 2002,298:1727-1728.

Montano MA, Nixon CP, Essex M. Dysregulation through the NF-kappaB enhancer and TATA box of the human immunodeficiency virus type 1 subtype E promoter. J Virol 1998,72:8446-8452.

Montavon C, Bibollet-Ruche F, Robertson D, et al. The identification of a complex A/G/I/J recombinant HIV type 1 virus in various West African countries. AIDS Res Hum Retroviruses 1999,15:1707-1712.

Montavon C, Toure-Kane C, Nkengasong JN, et al. - a CRF06-cpx: a new circulating recombinant form of HIV-1 in West Africa involving subtypes A, G, K, and J. J Acquir Immune Defic Syndr 2002,29:522-530.

Montavon C, Vergne L, Bourgeois A, et al. – b Identification of a new circulating recombinant form of HIV type 1, CRF11-cpx, involving subtypes A, G, J, and CRF01-AE, in Central Africa. AIDS Res Hum Retroviruses 2002,18:231-236.

Moore JP, Ho DD. HIV-1 neutralization: the consequences of viral adaptation to growth on transformed T cells. Aids 1995,9 Suppl A:S117-136.

Moore JP, Kitchen SG, Pugach P, Zack JA. The CCR5 and CXCR4 coreceptors--central to understanding the transmission and pathogenesis of human immunodeficiency virus type 1 infection. AIDS Res Hum Retroviruses 2004,20:111-126.

Morgan RW, Sofer L, Anderson AS, Bernberg EL, Cui J, Burnside J. Induction of host gene expression following infection of chicken embryo fibroblasts with oncogenic Marek's disease virus. J Virol 2001,75:533-539.

Morgenstern JP, Land H. Advanced mammalian gene transfer: high titre retroviral vectors with multiple drug selection markers and a complementary helper-free packaging cell line. Nucleic Acids Res 1990,18:3587-3596.

Morita E, Sundquist WI. Retrovirus budding. Annu Rev Cell Dev Biol 2004,20:395-425.

Moses AV, Jarvis MA, Raggo C, et al. A functional genomics approach to Kaposi's sarcoma. Ann N Y Acad Sci 2002,975:180-191.

Mossman KL, Macgregor PF, Rozmus JJ, Goryachev AB, Edwards AM, Smiley JR. Herpes simplex virus triggers and then disarms a host antiviral response. J Virol 2001,75:750-758.

Mulder LC, Muesing MA. Degradation of HIV-1 integrase by the N-end rule pathway. J Biol Chem 2000,275:29749-29753.

Mulders MN, van Loon AM, van der Avoort HG, et al. Molecular characterization of a wild poliovirus type 3 epidemic in The Netherlands (1992 and 1993). J Clin Microbiol 1995,33:3252-3256.

Murdaca J, Treins C, Monthouel-Kartmann MN, et al. Grb10 prevents Nedd4-mediated vascular endothelial growth factor receptor-2 degradation. J Biol Chem 2004,279:26754-26761.

Murphy G, Stanton H, Cowell S, et al. Mechanisms for pro matrix metalloproteinase activation. Apmis 1999,107:38-44.

Murray MC, Embree JE, Ramdahin SG, Anzala AO, Njenga S, Plummer FA. Effect of human immunodeficiency virus (HIV) type 1 viral genotype on mother-to-child transmission of HIV-1. J Infect Dis 2000,181:746-749.

Myers RE, Gale CV, Takeuchi Y, Harrison A, Kellam P. *in press* A Statistical Model for HIV-1 Sequence Classification using the Subtype Analyser (STAR). Bioinformatics.

Naderi A, Ahmed AA, Barbosa-Morais NL, Aparicio S, Brenton JD, Caldas C. Expression microarray reproducibility is improved by optimising purification steps in RNA amplification and labelling. BMC Genomics 2004,5:9.

Naghavi MH, Estable MC, Schwartz S, Roeder RG, Vahlne A. Upstream stimulating factor affects human immunodeficiency virus type 1 (HIV-1) long terminal repeat-directed transcription in a cell-specific manner, independently of the HIV-1 subtype and the core-negative regulatory element. J Gen Virol 2001,82:547-559.

Nasioulas G, Paraskevis D, Magiorkinis E, Theodoridou M, Hatzakis A. Molecular analysis of the full-length genome of HIV type 1 subtype I: evidence of A/G/I recombination. AIDS Res Hum Retroviruses 1999,15:745-758.

Natsumeda Y, Ohno S, Kawasaki H, Konno Y, Weber G, Suzuki K. Two distinct cDNAs for human IMP dehydrogenase. J Biol Chem 1990,265:5292-5295.

Nekhai S, Shukla RR, Fernandez A, Kumar A, Lamb NJ. Cell cycle-dependent stimulation of the HIV-1 promoter by Tat-associated CAK activator. Virology 2000,266:246-256.

Nelson JA, Fiscus SA, Swanstrom R. Evolutionary variants of the human immunodeficiency virus type 1 V3 region characterized by using a heteroduplex tracking assay. J Virol 1997,71:8750-8758.

Neoh LP, Akimoto H, Kaneko H, et al. The production of beta-chemokines induced by HIV-2 envelope glycoprotein. Aids 1997,11:1062-1063.

Nguyen DH, Hildreth JE. Evidence for budding of human immunodeficiency virus type 1 selectively from glycolipid-enriched membrane lipid rafts. J Virol 2000,74:3264-3272.

Nisole S, Saib A. Early steps of retrovirus replicative cycle. Retrovirology 2004,1:9.

Njouom R, Pasquier C, Sandres-Saune K, Harter A, Souyris C, Izopet J. Assessment of HIV-1 subtyping for Cameroon strains using phylogenetic analysis of pol gene sequences. J Virol Methods 2003,110:1-8.

Nottay BK, Kew OM, Hatch MH, Heyward JT, Obijeski JF. Molecular variation of type 1 vaccine-related and wild polioviruses during replication in humans. Virology 1981,108:405-423.

O'Brien TR, Blattner WA, Waters D, et al. Serum HIV-1 RNA levels and time to development of AIDS in the Multicenter Hemophilia Cohort Study. Jama 1996,276:105-110.

O'Brien TR, Rosenberg PS, Yellin F, Goedert JJ. Longitudinal HIV-1 RNA levels in a cohort of homosexual men. J Acquir Immune Defic Syndr Hum Retrovirol 1998,18:155-161.

O'Connell C, O'Brien S, Nash WG, Cohen M. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology 1984,138:225-235.

Oelrichs RB, Workman C, Laukkanen T, McCutchan FE, Deacon NJ. A novel subtype A/G/J recombinant full-length HIV type 1 genome from Burkina Faso. AIDS Res Hum Retroviruses 1998,14:1495-1500.

Okabe H, Satoh S, Kato T, et al. Genome-wide analysis of gene expression in human hepatocellular carcinomas using cDNA microarray: identification of genes involved in viral carcinogenesis and tumor progression. Cancer Res 2001,61:2129-2137.

Ono A, Freed EO. Plasma membrane rafts play a critical role in HIV-1 assembly and release. Proc Natl Acad Sci U S A 2001,98:13925-13930.

Ott DE. Cellular proteins in HIV virions. Rev Med Virol 1997,7:167-180.

Ott DE. Potential roles of cellular proteins in HIV-1. Rev Med Virol 2002,12:359-374.

Ott DE, Coren LV, Johnson DG, Sowder RC, 2nd, Arthur LO, Henderson LE. Analysis and localization of cyclophilin A found in the virions of human immunodeficiency virus type 1 MN strain. AIDS Res Hum Retroviruses 1995,11:1003-1006.

Ott DE, Coren LV, Kane BP, et al. Cytoskeletal proteins inside human immunodeficiency virus type 1 virions. J Virol 1996,70:7734-7743.

Ou CY, Ciesielski CA, Myers G, et al. Molecular epidemiology of HIV transmission in a dental practice. Science 1992,256:1165-1171.

Oxtoby MJ. Perinatally acquired human immunodeficiency virus infection. Pediatr Infect Dis J 1990,9:609-619.

Palacios EH, Weiss A. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. Oncogene 2004,23:7990-8000.

Palma JP, Kim BS. The scope and activation mechanisms of chemokine gene expression in primary astrocytes following infection with Theiler's virus. J Neuroimmunol 2004,149:121-129.

Palmiter RD. The elusive function of metallothioneins. Proc Natl Acad Sci U S A 1998,95:8428-8430.

Pantaleo G, Fauci AS. Immunopathogenesis of HIV infection. Annu Rev Microbiol 1996,50:825-854.

Paraskevis D, Magiorkinis E, Magiorkinis G, et al. Molecular characterization of a complex, recombinant human immunodeficiency virus type 1 (HIV-1) isolate (A/G/J/K/?): evidence to support the existence of a novel HIV-1 subtype. J Gen Virol 2001,82:2509-2514.

Paraskevis D, Magiorkinis M, Paparizos V, Pavlakis GN, Hatzakis A. Molecular characterization of a recombinant HIV type 1 isolate (A/G/E/?): unidentified regions may be derived from parental subtype E sequences. AIDS Res Hum Retroviruses 2000,16:845-855.

Parekh B, Phillips S, Granade TC, Baggs J, Hu DJ, Respess R. Impact of HIV type 1 subtype variation on viral RNA quantitation. AIDS Res Hum Retroviruses 1999,15:133-142.

Parikh U, Clark, SA, Calef C and Mellors, JA. Mutations in Retroviral Genes Associated with Drug Resistance 2001. In: HIV Sequence Compendium. Edited by Leitner T FB, Hahn B, Marx P, McCutchan F, Mellors J, Wolinsky S, and Korber B.: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, LA-UR number 04-7420; 2001:38-122.

Parry JV, Murphy G, Barlow KL, et al. National surveillance of HIV-1 subtypes for England and Wales: design, methods, and initial findings. J Acquir Immune Defic Syndr 2001,26:381-388.

Pasquier C, Millot N, Njouom R, et al. HIV-1 subtyping using phylogenetic analysis of pol gene sequences. J Virol Methods 2001,94:45-54.

Peeters M. The genetic variability of HIV-1 and its implications. Transfus Clin Biol 2001,8:222-225.

Pereira LA, Bentley K, Peeters A, Churchill MJ, Deacon NJ. A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter. Nucleic Acids Res 2000,28:663-668.

Perez-Alvarez L, Cuevas MT, Villahermosa ML, et al. Prevalence of drug resistance mutations in B, non-B subtypes, and recombinant forms of human immunodeficiency virus type 1 in infected individuals in Spain (Galicia). J Hum Virol 2001,4:35-38.

Perez-Alvarez L, Thomson MM, Villahermosa ML, et al. HIV-1 subtype G and BG recombinant viruses in Spanish natives: evidence of characteristic mutations in reverse transcriptase and protease. Aids 2001,15:1907-1910.

Perou CM. Show me the data! Nat Genet 2001,29:373.

Peruzzi F, Gordon J, Darbinian N, Amini S. Tat-induced deregulation of neuronal differentiation and survival by nerve growth factor pathway. J Neurovirol 2002,8 Suppl 2:91-96.

Petropoulos CJ, Parkin NT, Limoli KL, et al. A Novel Phenotypic Drug Susceptibility Assay for Human Immunodeficiency Virus Type 1. Antimicrob. Agents Chemother. 2000,44:920-928.

Pezzotti P, Phillips AN, Dorrucci M, et al. Category of exposure to HIV and age in the progression to AIDS: longitudinal study of 1199 people with known dates of seroconversion. HIV Italian Seroconversion Study Group. Bmj 1996,313:583-586.

Pietiainen V, Huttunen P, Hyypia T. Effects of echovirus 1 infection on cellular gene expression. Virology 2000,276:243-250.

Pillai S, Good B, Richman D, Corbeil J. A new perspective on V3 phenotype prediction. AIDS Res Hum Retroviruses 2003,19:145-149.

Pillay D, Taylor S, Richman DD. Incidence and impact of resistance against approved antiretroviral drugs. Rev Med Virol 2000,10:231-253.

Pillay D, Walker AS, Gibb DM, et al. Impact of human immunodeficiency virus type 1 subtypes on virologic response and emergence of drug resistance among children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 trial. J Infect Dis 2002,186:617-625.

Pirounaki M, Heyden NA, Arens M, Ratner L. Rapid phenotypic drug susceptibility assay for HIV-1 with a CCR5 expressing indicator cell line. J Virol Methods 2000,85:151-161.

Piyasirisilp S, McCutchan FE, Carr JK, et al. A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. J Virol 2000,74:11286-11295.

Planelles V, Jowett JB, Li QX, Xie Y, Hahn B, Chen IS. Vpr-induced cell cycle arrest is conserved among primate lentiviruses. J Virol 1996,70:2516-2524.

Poggioli GJ, DeBiasi RL, Bickel R, et al. Reovirus-induced alterations in gene expression related to cell cycle regulation. J Virol 2002,76:2585-2594.

Pohlmann S, Baribaud F, Doms RW. DC-SIGN and DC-SIGNR: helping hands for HIV. Trends Immunol 2001,22:643-646.

Pollakis G, Kang S, Kliphuis A, Chalaby MI, Goudsmit J, Paxton WA. N-linked glycosylation of the HIV type-1 gp120 envelope glycoprotein as a major determinant of CCR5 and CXCR4 coreceptor utilization. J Biol Chem 2001,276:13433-13441.

Poole LJ, Yu Y, Kim PS, Zheng QZ, Pevsner J, Hayward GS. Altered patterns of cellular gene expression in dermal microvascular endothelial cells infected with Kaposi's sarcoma-associated herpesvirus. J Virol 2002,76:3395-3420.

Poon B, Chen IS. Human immunodeficiency virus type 1 (HIV-1) Vpr enhances expression from unintegrated HIV-1 DNA. J Virol 2003,77:3962-3972.

Pope M, Frankel SS, Mascola JR, et al. Human immunodeficiency virus type 1 strains of subtypes B and E replicate in cutaneous dendritic cell-T-cell mixtures without displaying subtype-specific tropism. J Virol 1997,71:8001-8007.

Popov S, Rexach M, Ratner L, Blobel G, Bukrinsky M. - a Viral protein R regulates docking of the HIV-1 preintegration complex to the nuclear pore complex. J Biol Chem 1998,273:13347-13352.

Popov S, Rexach M, Zybarth G, et al. - b Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. Embo J 1998,17:909-917.

Sarngadharan MG, Read E, Gallo RC. Detection, isolation, and continuous production of cytopathic retroviruses (HTLV-III) from patients with AIDS and pre-AIDS. Science 1984,224:497-500.

Popper SJ, Sarr AD, Gueye-Ndiaye A, Mboup S, Essex ME, Kanki PJ. Low plasma human immunodeficiency virus type 2 viral load is independent of proviral load: low virus production in vivo. J Virol 2000,74:1554-1557.

Popper SJ, Sarr AD, Travers KU, et al. Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. J Infect Dis 1999,180:1116-1121.

Poulsen AG, Kvinesdal B, Aaby P, et al. Prevalence of and mortality from human immunodeficiency virus type 2 in Bissau, West Africa. Lancet 1989,1:827-831.

Py B, Slomianny C, Auberger P, Petit PX, Benichou S. Siva-1 and an alternative splice form lacking the death domain, Siva-2, similarly induce apoptosis in T lymphocytes via a caspase-dependent mitochondrial pathway. J Immunol 2004,172:4008-4017.

Quinones-Mateu ME, Ball SC, Marozsan AJ, et al. A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. J Virol 2000,74:9222-9233.

Quinones-Mateu ME, Tadele M, Parera M, et al. Insertions in the reverse transcriptase increase both drug resistance and viral fitness in a human immunodeficiency virus type 1 isolate harboring the multi-nucleoside reverse transcriptase inhibitor resistance 69 insertion complex mutation. J Virol 2002,76:10546-10552.

Quivy V, Adam E, Collette Y, et al. Synergistic activation of human immunodeficiency virus type 1 promoter activity by NF-kappaB and inhibitors of deacetylases: potential perspectives for the development of therapeutic strategies. J Virol 2002,76:11091-11103.

Raab-Traub N, Flynn K. The structure of the termini of the Epstein-Barr virus as a marker of clonal cellular proliferation. Cell 1986,47:883-889.

Rambaut A, Robertson DL, Pybus OG, Peeters M, Holmes EC. Human immunodeficiency virus. Phylogeny and the origin of HIV-1. Nature 2001,410:1047-1048.

Ramsay G. DNA chips: state-of-the art. Nat Biotechnol 1998,16:40-44.

Rathod PK, Ganesan K, Hayward RE, Bozdech Z, DeRisi JL. DNA microarrays for malaria. Trends Parasitol 2002,18:39-45.

Ratner L, Haseltine W, Patarca R, et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 1985,313:277-284.

Raulin J. Human immunodeficiency virus and host cell lipids. Interesting pathways in research for a new HIV therapy. Progress in Lipid Research 2002,41:27-65.

Ray N, Enquist LW. Transcriptional response of a common permissive cell type to infection by two diverse alphaherpesviruses. J Virol 2004,78:3489-3501.

Reeves JD, Doms RW. Human immunodeficiency virus type 2. J Gen Virol 2002,83:1253-1265.

Reil H, Bukovsky AA, Gelderblom HR, Gottlinger HG. Efficient HIV-1 replication can occur in the absence of the viral matrix protein. Embo J 1998,17:2699-2708.

Renjifo B, Fawzi W, Mwakagile D, et al. Differences in perinatal transmission among human immunodeficiency virus type 1 genotypes. J Hum Virol 2001,4:16-25.

Renjufi B, Gilbert P, Chaplin B, et al. Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. Aids 2004,18:1629-1636.

Renwick N, Weverling GJ, Brouwer J, Bakker M, Schulz TF, Goudsmit J. Vascular endothelial growth factor levels in serum do not increase following HIV type 1 and HHV8 seroconversion and lack correlation with AIDS-related Kaposi's sarcoma. AIDS Res Hum Retroviruses 2002,18:695-698.

Richman DD, Wrin T, Little SJ, Petropoulos CJ. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 2003,100:4144-4149.

Richmond CS, Glasner JD, Mau R, Jin H, Blattner FR. Genome-wide expression profiling in Escherichia coli K-12. Nucleic Acids Res 1999,27:3821-3835.

Rinke de Wit TF, Tsegaye A, Wolday D, et al. Primary HIV-1 subtype C infection in Ethiopia. J Acquir Immune Defic Syndr 2002,30:463-470.

Rivera VM, Welsh JD, Maizel JV, Jr. Comparative sequence analysis of the 5' noncoding region of the enteroviruses and rhinoviruses. Virology 1988,165:42-50.

Roberts ES, Zandonatti MA, Watry DD, et al. Induction of pathogenic sets of genes in macrophages and neurons in NeuroAIDS. Am J Pathol 2003,162:2041-2057.

Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. Science 2000,288:55-56.

Rodes B, Toro C, Paxinos E, et al. Differences in disease progression in a cohort of long-term non-progressors after more than 16 years of HIV-1 infection. Aids 2004,18:1109-1116.

Rogel ME, Wu LI, Emerman M. The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection. J Virol 1995,69:882-888.

Romani N, Ratzinger G, Pfaller K, et al. Migration of dendritic cells into lymphatics-the Langerhans cell example: routes, regulation, and relevance. Int Rev Cytol 2001,207:237-270.

Rose JJ, Janvier K, Chandrasekhar S, Sekaly RP, Bonifacino JS, Venkatesan S. CD4 downregulation by HIV-1 and SIV Nef proteins involves both internalization and intracellular retention mechanisms. J Biol Chem 2004. Dec 16; Epub ahead of print

Rosenberger CM, Scott MG, Gold MR, Hancock RE, Finlay BB. Salmonella typhimurium infection and lipopolysaccharide stimulation induce similar changes in macrophage gene expression. J Immunol 2000,164:5894-5904.

Rosu-Myles M, Gallacher L, Murdoch B, et al. The human hematopoietic stem cell compartment is heterogeneous for CXCR4 expression. Proc Natl Acad Sci U S A 2000,97:14626-14631.

Rozanov M, Plikat U, Chappey C, Kochergin A, Tatusova T. A web-based genotyping resource for viral sequences. Nucl. Acids Res. 2004,32:W654-659.

Rubsamen-Waigmann H, von Briesen H, Holmes H, et al. Standard conditions of virus isolation reveal biological variability of HIV type 1 in different regions of the world. WHO Network for HIV Isolation and Characterization. AIDS Res Hum Retroviruses 1994,10:1401-1408.

Ruibal-Brunet IJ, Cuevas MT, Diaz-Torres H, et al. Genotypic resistance mutations to antiretroviral drugs in HIV-1 B and non-B subtypes from Cuba. Rev Panam Salud Publica 2001,10:174-180.

Safai B, Sarngadharan MG, Groopman JE, et al. Seroepidemiological studies of human T-lymphotropic retrovirus type III in acquired immunodeficiency syndrome. Lancet 1984,1:1438-1440.

Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987,4:406-425.

Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. A whole-genome microarray reveals genetic diversity among Helicobacter pylori strains. Proc Natl Acad Sci U S A 2000,97:14668-14673.

Samson M, Libert F, Doranz BJ, et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. Nature 1996,382:722-725.

Sanger F, Air GM, Barrell BG, et al. Nucliotide sequence of bacteriophage phi X174 DNA. Nature 1977,265:687-695.

Sarngadharan MG, DeVico AL, Bruch L, Schupbach J, Gallo RC. HTLV-III: the etiologic agent of AIDS. Princess Takamatsu Symp 1984,15:301-308.

Sattentau QJ, Weiss RA. The CD4 antigen: physiological ligand and HIV receptor. Cell 1988,52:631-633.

Schacker TW, Hughes JP, Shea T, Coombs RW, Corey L. Biological and virologic characteristics of primary HIV infection. Ann Intern Med 1998,128:613-620.

Schaeffer E, Soros VB, Greene WC. Compensatory link between fusion and endocytosis of human immunodeficiency virus type 1 in human CD4 T lymphocytes. J Virol 2004,78:1375-1383.

Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995,270:467-470.

Scheurer D. Rapid reversal of renal failure after initiation of HAART: a case report. AIDS Read 2004,14:443-447.

Schneider J, Buness A, Huber W, et al. Systematic analysis of T7 RNA polymerase based in vitro linear RNA amplification for use in microarray experiments. BMC Genomics 2004,5:29.

Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 2002,110:521-529.

Schwartz S, Felber BK, Benko DM, Fenyo EM, Pavlakis GN. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J Virol 1990,64:2519-2529.

Schwarz A, Offermann G, Keller F, et al. The effect of cyclosporine on the progression of human immunodeficiency virus type 1 infection transmitted by transplantation--data on four cases and review of the literature. Transplantation 1993,55:95-103.

Seeger M, Ferrell K, Frank R, Dubiel W. HIV-1 tat inhibits the 20 S proteasome and its 11 S regulator-mediated activation. J Biol Chem 1997,272:8145-8148.

Shaffer N, Roongpisuthipong A, Siriwasin W, et al. Maternal virus load and perinatal human immunodeficiency virus type 1 subtype E transmission, Thailand. Bangkok Collaborative Perinatal HIV Transmission Study Group. J Infect Dis 1999,179:590-599.

Shaheduzzaman S, Krishnan V, Petrovic A, et al. Effects of HIV-1 Nef on cellular gene expression profiles. J Biomed Sci 2002,9:82-96.

Shea A, Sarr DA, Jones N, et al. CCR5 receptor expression is down-regulated in HIV type 2 infection: implication for viral control and protection. AIDS Res Hum Retroviruses 2004,20:630-635.

Shearer GM, Clerici M. Cytokine profiles in HIV type 1 disease and protection. AIDS Res Hum Retroviruses 1998,14 Suppl 2:S149-152.

Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. Nature 2002,418:646-650.

Sheehy AM, Gaddis NC, Malim MH. The antiretroviral enzyme APOBEC3G is degraded by the proteasome in response to HIV-1 Vif. Nat Med 2003,9:1404-1407.

Sheppard HW, Lang W, Ascher MS, Vittinghoff E, Winkelstein W. The characterization of non-progressors: long-term HIV-1 infection with stable CD4+ T-cell levels. Aids 1993,7:1159-1166.

Sherer NM, Lehmann MJ, Jimenez-Soto LF, et al. Visualization of Retroviral Replication in Living Cells Reveals Budding into Multivesicular Bodies. Traffic 2003,4:785-801.

Sherman MP, Greene WC. Slipping through the door: HIV entry into the nucleus. Microbes Infect 2002,4:67-73.

Sidagis J, Ueno K, Tokunaga M, Ohyama M, Eizuru Y. Molecular epidemiology of Epstein-Barr virus (EBV) in EBV-related malignancies. Int J Cancer 1997,72:72-76.

Simmen KA, Singh J, Luukkonen BG, et al. Global modulation of cellular transcription by human cytomegalovirus is initiated by viral glycoprotein B. Proc Natl Acad Sci U S A 2001,98:7140-7145.

Simmons A, Aluvihare V, McMichael A. Nef triggers a transcriptional program in T cells imitating single-signal T cell activation and inducing HIV virulence mediators. Immunity 2001,14:763-777.

Simmons G, Reeves JD, McKnight A, et al. CXCR4 as a functional coreceptor for human immunodeficiency virus type 1 infection of primary macrophages. J Virol 1998,72:8453-8457.

Simon F, Ly TD, Baillou-Beaufils A, et al. Sensitivity of screening kits for anti-HIV-1 subtype O antibodies. Aids 1994,8:1628-1629.

Simon F, Mauclere P, Roques P, et al. Identification of a new human immunodeficiency virus type 1 distinct from group M and group O. Nat Med 1998,4:1032-1037.

Smith PD, Meng G, Salazar-Gonzalez JF, Shaw GM. Macrophage HIV-1 infection and the gastrointestinal tract reservoir. J Leukoc Biol 2003,74:642-649.

Smith TF, Srinivasan A, Schochetman G, Marcus M, Myers G. The phylogenetic history of immunodeficiency viruses. Nature 1988,333:573-575.

Soda Y, Shimizu N, Jinno A, et al. Establishment of a new system for determination of coreceptor usages of HIV based on the human glioma NP-2 cell line. Biochem Biophys Res Commun 1999,258:313-321.

Sonigo P, Alizon M, Staskus K, et al. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell 1985,42:369-382.

Soto-Ramirez LE, Renjifo B, McLane MF, et al. HIV-1 Langerhans' cell tropism associated with heterosexual transmission of HIV. Science 1996,271:1291-1293.

Southern E, Mir K, Shchepinov M. Molecular interactions on microarrays. Nat Genet 1999,21:5-9.

Southern EM. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 1975,98:503-517.

Spector DH, Hock L, Tamashiro JC. Cleavage maps for human cytomegalovirus DNA strain AD169 for restriction endonucleases EcoRI, BglII, and HindIII. J Virol 1982,42:558-582.

Spenlehauer C, Gordon CA, Trkola A, Moore JP. A luciferase-reporter gene-expressing T-cell line facilitates neutralization and drug-sensitivity assays that use either R5 or X4 strains of human immunodeficiency virus type 1. Virology 2001,280:292-300.

Spiess AN, Mueller N, Ivell R. Amplified RNA degradation in T7-amplification methods results in biased microarray hybridizations. BMC Genomics 2003,4:44.

Sporer B, Koedel U, Paul R, Eberle J, Arendt G, Pfister HW. Vascular endothelial growth factor (VEGF) is increased in serum, but not in cerebrospinal fluid in HIV associated CNS diseases. J Neurol Neurosurg Psychiatry 2004,75:298-300.

Steinman L. Immune therapy for autoimmune diseases. Science 2004,305:212-216.

Stevenson M, Haggerty S, Lamonica CA, Meier CM, Welch SK, Wasiak AJ. Integration is not necessary for expression of human immunodeficiency virus type 1 protein products. J Virol 1990,64:2421-2425.

Stingley SW, Ramirez JJ, Aguilar SA, et al. Global analysis of herpes simplex virus type 1 transcription using an oligonucleotide-based DNA microarray. J Virol 2000,74:9916-9927.

Stoffel A, Levine AJ. Actvation of NF-kappaB by the API2/MALT1 fusions inhibits p53 dependant but not FAS induced apoptosis: a directional link between NF-kappaB and p53. Cell Cycle 2004,3:1017-1020.

Stolk LM, Luers JF. Increasing number of anti-HIV drugs but no definite cure. Review of anti-HIV drugs. Pharm World Sci 2004,26:133-136.

Stopak K, de Noronha C, Yonemoto W, Greene WC. HIV-1 Vif blocks the antiviral activity of APOBEC3G by impairing both its translation and intracellular stability. Mol Cell 2003,12:591-601.

Storch GA, Anderson LJ, Park CS, Tsou C, Dohner DE. Antigenic and genomic diversity within group A respiratory syncytial virus. J Infect Dis 1991,163:858-861.

Studier JA, Keppler KJ. A note on the neighbor-joining algorithm of Saitou and Nei. Mol Biol Evol 1988,5:729-731.

Subbarao K, Klimov A, Katz J, et al. Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. Science 1998,279:393-396.

Subbarao S, Vanichseni S, Hu DJ, et al. Genetic characterization of incident HIV type 1 subtype E and B strains from a prospective cohort of injecting drug users in Bangkok, Thailand. AIDS Res Hum Retroviruses 2000,16:699-707.

Takayama M, Takayama N, Inoue N, Kameoka Y. Application of long PCR method of identification of variations in nucleotide sequences among varicella-zoster virus isolates. J Clin Microbiol 1996,34:2869-2874.

Talaat AM, Hunter P, Johnston SA. Genome-directed primers for selective labeling of bacterial transcripts for DNA microarray analysis. Nat Biotechnol 2000,18:679-682.

Tamayo P, Slonim D, Mesirov J, et al. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci U S A 1999,96:2907-2912.

Tang H, Kuhen KL, Wong-Staal F. Lentivirus replication and regulation. Annu Rev Genet 1999,33:133-170.

Tapia N, Franco S, Puig-Basagoiti F, et al. Influence of human immunodeficiency virus type 1 subtype on mother-to-child transmission. J Gen Virol 2003,84:607-613.

Tatt ID, Barlow KL, Clewley JP. A gag gene heteroduplex mobility assay for subtyping HIV-1. J Virol Methods 2000,87:41-51.

Tatt ID, Barlow KL, Clewley JP, Gill ON, Parry JV. Surveillance of HIV-1 Subtypes Among Heterosexuals in England and Wales, 1997-2000. J Acquir Immune Defic Syndr 2004,36:1092-1099.

Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res 2000,28:33-36.

Tau G, Rothman P. Biologic functions of the IFN-gamma receptors. Allergy 1999,54:1233-1251.

Taylor LA, Carthy CM, Yang D, et al. Host gene regulation during coxsackievirus B3 infection in mice: assessment by microarrays. Circ Res 2000,87:328-334.

Tchou I, Misery L, Sabido O, et al. Functional HIV CXCR4 coreceptor on human epithelial Langerhans cells and infection by HIV strain X4. J Leukoc Biol 2001,70:313-321.

Teo I, Veryard C, Barnes H, et al. Circular forms of unintegrated human immunodeficiency virus type 1 DNA and high levels of viral protein expression: association with dementia and multinucleated giant cells in the brains of patients with AIDS. J Virol 1997,71:2928-2933.

Tersmette M, de Goede RE, Al BJ, et al. Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. J Virol 1988,62:2026-2032.

Tersmette M, Lange JM, de Goede RE, et al. Association between biological properties of human immunodeficiency virus variants and risk for AIDS and AIDS mortality. Lancet 1989,1:983-985.

Thali M, Bukovsky A, Kondo E, et al. Functional association of cyclophilin A with HIV-1 virions. Nature 1994,372:363-365.

The HIV Infection in Newborns French Collaborative Study Group. Comparison of vertical human immunodeficiency virus type 2 and human immunodeficiency virus type 1 transmission in the French prospective cohort. Pediatr Infect Dis J. 1994,13:502-506.

The Washington Blade Gay cancer focus of hearing. 1982.

Thomas JG, Olson JM, Tapscott SJ, Zhao LP. An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. Genome Res 2001,11:1227-1236.

Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 1994,22:4673-4680.

Thomson MM, Delgado E, Herrero I, et al. Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. J Gen Virol 2002,83:107-119.

Tian B, Zhang Y, Luxon BA, et al. Identification of NF-kappaB-dependent gene networks in respiratory syncytial virus-infected cells. J Virol 2002,76:6800-6814.

Tian Q, Taupin J, Elledge S, Robertson M, Anderson P. Fas-activated serine/threonine kinase (FAST) phosphorylates TIA-1 during Fas-mediated apoptosis. J Exp Med 1995,182:865-874.

Tolpin MD, Stewart JA, Warren D, et al. Transfusion transmission of cytomegalovirus confirmed by restriction endonuclease analysis. J Pediatr 1985,107:953-956.

Toschi E, Barillari G, Sgadari C, et al. Activation of matrix-metalloproteinase-2 and membrane-type-1-matrix-metalloproteinase in endothelial cells and induction of vascular permeability in vivo by human immunodeficiency virus-1 Tat protein and basic fibroblast growth factor. Mol Biol Cell 2001,12:2934-2946.

Tovanabutra S, Watanaveeradej V, Viputtikul K, et al. A new circulating recombinant form, CRF15_01B, reinforces the linkage between IDU and heterosexual epidemics in Thailand. AIDS Res Hum Retroviruses 2003,19:561-567.

Towers GJ, Hatziioannou T, Cowan S, Goff SP, Luban J, Bieniasz PD. Cyclophilin A modulates the sensitivity of HIV-1 to host restriction factors. Nat Med 2003,9:1138-1143.

Travers SA, O'Connell MJ, McCormack GP, McInerney JO. Evidence for heterogeneous selective pressures in the evolution of the env gene in different human immunodeficiency virus type 1 subtypes. J Virol 2005,79:1836-1841.

Triques K, Bourgeois A, Saragosti S, et al. High diversity of HIV-1 subtype F strains in Central Africa. Virology 1999,259:99-109.

Triques K, Bourgeois A, Vidal N, et al. Near-full-length genome sequencing of divergent African HIV type 1 subtype F viruses leads to the identification of a new HIV type 1 subtype designated K. AIDS Res Hum Retroviruses 2000,16:139-151.

Trkola A, Matthews J, Gordon C, Ketas T, Moore JP. A Cell Line-Based Neutralization Assay for Primary Human Immunodeficiency Virus Type 1 Isolates That Use either the CCR5 or the CXCR4 Coreceptor. J. Virol. 1999,73:8966-8974.

Trono D. Partial reverse transcripts in virions from human immunodeficiency and murine leukemia viruses. J Virol 1992,66:4893-4900.

Tsavachidou D, Podrzucki W, Seykora J, Berger SL. Gene array analysis reveals changes in peripheral nervous system gene expression following stimuli that result in reactivation of latent herpes simplex virus type 1: induction of transcription factor Bcl-3. J Virol 2001,75:9909-9917.

Tscherning C, Alaeus A, Fredriksson R, et al. Differences in chemokine coreceptor usage between genetic subtypes of HIV-1. Virology 1998,241:181-188.

Tseng GC, Oh MK, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic Acids Res 2001,29:2549-2557.

Tuaillon E, Gueudin M, Lemee V, et al. Phenotypic Susceptibility to Nonnucleoside Inhibitors of Virion-Associated Reverse Transcriptase From Different HIV Types and Groups. J Acquir Immune Defic Syndr 2004,37:1543-1549.

Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001,98:5116-5121.

Vahey MT, Nau ME, Jagodzinski LL, et al. Impact of viral infection on the gene expression profiles of proliferating normal human peripheral blood mononuclear cells infected with HIV type 1 RF. AIDS Res Hum Retroviruses 2002,18:179-192.

Valentin A, Trivedi H, Lu W, Kostrikis LG, Pavlakis GN. CXCR4 mediates entry and productive infection of syncytia-inducing (X4) HIV-1 strains in primary macrophages. Virology 2000,269:294-304.

Van der Auwera G, Janssens W, Heyndrickx L, van der Groen G. Reanalysis of full-length HIV type 1 group M subtype K and sub-subtype F2 with an MS-DOS bootscanning program. AIDS Res Hum Retroviruses 2001,17:185-189.

Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD, Eberwine JH. Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A 1990,87:1663-1667.

van Opijnen T, Jeeninga RE, Boerlijst MC, et al. Human immunodeficiency virus type 1 subtypes have a distinct long terminal repeat that determines the replication rate in a host-cell-specific manner. J Virol 2004,78:3675-3683.

van 't Wout AB, Lehrman GK, Mikheeva SA, et al. Cellular gene expression upon human immunodeficiency virus type 1 infection of CD4(+)-T-cell lines. J Virol 2003,77:1392-1402.

Varin A, Manna SK, Quivy V, et al. Exogenous Nef protein activates NF-kappa B, AP-1, and c-Jun N-terminal kinase and stimulates HIV transcription in promonocytic cells. Role in AIDS pathogenesis. J Biol Chem 2003,278:2219-2227.

Vartanian JP, Meyerhans A, Asjo B, Wain-Hobson S. Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. J Virol 1991,65:1779-1788.

Velazquez-Campoy A, Todd MJ, Vega S, Freire E. Catalytic efficiency and vitality of HIV-1 proteases from African viral subtypes. Proc Natl Acad Sci U S A 2001,98:6062-6067.

Vendeville A, Rayne F, Bonhoure A, Bettache N, Montcourrier P, Beaumelle B. HIV-1 Tat Enters T Cells Using Coated Pits before Translocating from Acidified Endosomes and Eliciting Biological Responses. Mol. Biol. Cell 2004,15:2347-2360.

Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. Science 2001,291:1304-1351.

Verani A, Gras G, Pancino G. Macrophages and HIV-1: dangerous liaisons. Mol Immunol 2005,42:195-212.

Vergne L, Malonga-Mouellet G, Mistoul I, et al. Resistance to antiretroviral treatment in Gabon: need for implementation of guidelines on antiretroviral therapy use and HIV-1 drug resistance monitoring in developing countries. J Acquir Immune Defic Syndr 2002,29:165-168.

Vergne L, Peeters M, Mpoudi-Ngole E, et al. Genetic diversity of protease and reverse transcriptase sequences in non-subtype-B human immunodeficiency virus type 1 strains: evidence of many minor drug resistance mutations in treatment-naive patients. J Clin Microbiol 2000,38:3919-3925.

Vertegaal AC, Kuiperij HB, Houweling A, Verlaan M, van der Eb AJ, Zantema A. Differential expression of tapasin and immunoproteasome subunits in adenovirus type 5- versus type 12-transformed cells. J Biol Chem 2003,278:139-146.

Vertegaal AC, Kuiperij HB, van Laar T, Scharnhorst V, van der Eb AJ, Zantema A. cDNA micro array identification of a gene differentially expressed in adenovirus type 5- versus type 12-transformed cells. FEBS Lett 2000,487:151-155.

Vidal N, Mulanga-Kabeya C, Nzilambi N, Delaporte E, Peeters M. - a Identification of a complex env subtype E HIV type 1 virus from the democratic republic of congo, recombinant with A, G, H, J, K, and unknown subtypes. AIDS Res Hum Retroviruses 2000,16:2059-2064.

Vidal N, Peeters M, Mulanga-Kabeya C, et al. – b Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. J Virol 2000,74:10498-10507.

Vlahov D, Graham N, Hoover D, et al. Prognostic indicators for AIDS and infectious disease death in HIV-infected injection drug users: plasma viral load and CD4+ cell count. Jama 1998,279:35-40.

von Schwedler U, Song J, Aiken C, Trono D. Vif is crucial for human immunodeficiency virus type 1 proviral DNA synthesis in infected cells. J Virol 1993,67:4945-4955.

Wagner A, Doerks A, Aboud M, et al. Induction of cellular genes is mediated by the Bel1 transactivator in foamy virus-infected human cells. J Virol 2000,74:4441-4447.

Wainberg MA. HIV-1 subtype distribution and the problem of drug resistance. Aids 2004,18 Suppl 3:S63-68.

Wain-Hobson S, Sonigo P, Danos O, Cole S, Alizon M. Nucleotide sequence of the AIDS virus, LAV. Cell 1985,40:9-17.

Wang HW, Trotter MW, Lagos D, et al. Kaposi sarcoma herpesvirus-induced cellular reprogramming contributes to the lymphatic endothelial gene expression in Kaposi sarcoma. Nat Genet 2004,36:687-693.

Wang L, Zhang H, Solski PA, Hart MJ, Der CJ, Su L. Modulation of HIV-1 replication by a novel RhoA effector activity. J Immunol 2000,164:5369-5374.

Wang T, Weinman SA. Involvement of chloride channels in hepatic copper metabolism: ClC-4 promotes copper incorporation into ceruloplasmin. Gastroenterology 2004,126:1157-1166.

Warke RV, Xhaja K, Martin KJ, et al. Dengue virus induces novel changes in gene expression of human umbilical vein endothelial cells. J Virol 2003,77:11822-11832.

Wasi C, Herring B, Raktham S, et al. Determination of HIV-1 subtypes in injecting drug users in Bangkok, Thailand, using peptide-binding enzyme immunoassay and heteroduplex mobility assay: evidence of increasing infection with HIV-1 subtype E. Aids 1995,9:843-849.

Watson A, Mazumder A, Stewart M, Balasubramanian S. Technology for microarray analysis of gene expression. Curr Opin Biotechnol 1998,9:609-614.

Weidle PJ, Malamba S, Mwebaze R, et al. Assessment of a pilot antiretroviral drug therapy programme in Uganda: patients' response, survival, and drug resistance. Lancet 2002,360:34-40.

Weinberg JB, Matthews TJ, Cullen BR, Malim MH. Productive human immunodeficiency virus type 1 (HIV-1) infection of nonproliferating human monocytes. J Exp Med 1991,174:1477-1482.

Weller I, Crawford DH, Iliescu V, et al. Homosexual men in London: lymphadenopathy, immune status, and Epstein-Barr virus infection. Ann N Y Acad Sci 1984,437:238-253.

Wendler WM, Kremmer E, Forster R, Winnacker EL. Identification of pirin, a novel highly conserved nuclear protein. J Biol Chem 1997,272:8482-8489.

WHO. Acquired Immune Deficiency Syndrome Emergencies. WHO Meeting. Geneva, 22-25, November 1983.

Wilbe K, Casper C, Albert J, Leitner T. Identification of two CRF11-cpx genomes and two preliminary representatives of a new circulating recombinant form (CRF13-cpx) of HIV type 1 in Cameroon. AIDS Res Hum Retroviruses 2002,18:849-856.

Willey SJ, Reeves JD, Hudson R, et al. Identification of a subset of human immunodeficiency virus type 1 (HIV-1), HIV-2, and simian immunodeficiency virus strains able to exploit an alternative coreceptor on untransformed human brain and lymphoid cells. J Virol 2003,77:6138-6152.

Wilson CL, Pepper SD, Hey Y, Miller CJ. Amplification protocols introduce systematic but reproducible errors into gene expression studies. Biotechniques 2004,36:498-506.

Wilson M, DeRisi J, Kristensen HH, et al. Exploring drug-induced alterations in gene expression in Mycobacterium tuberculosis by microarray hybridization. Proc Natl Acad Sci U S A 1999,96:12833-12838.

Winston DJ, Huang ES, Miller MJ, et al. Molecular epidemiology of cytomegalovirus infections associated with bone marrow transplantation. Ann Intern Med 1985,102:16-20.
Wolfinger RD, Gibson G, Wolfinger ED, et al. Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol 2001,8:625-637.

Won J, Kim DY, La M, Kim D, Meadows GG, Joe CO. Cleavage of 14-3-3 protein by caspase-3 facilitates bad interaction with Bcl-x(L) during apoptosis. J Biol Chem 2003,278:19347-19351.

Wu CG, Salvay DM, Forgues M, et al. Distinctive gene expression profiles associated with Hepatitis B virus x protein. Oncogene 2001,20:3674-3682.

Wu KJ, Polack A, Dalla-Favera R. Coordinated regulation of iron-controlling genes, H-ferritin and IRP2, by c-MYC. Science 1999,283:676-679.

Wu Y, Marsh JW. Early transcription from nonintegrated DNA in human immunodeficiency virus infection. J Virol 2003,77:10376-10382.

Xiao H, Tao Y, Greenblatt J, Roeder RG. A cofactor, TIP30, specifically enhances HIV-1 Tat-activated transcription. Proc Natl Acad Sci U S A 1998,95:2146-2151.

Xu XR, Huang J, Xu ZG, et al. Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. Proc Natl Acad Sci U S A 2001,98:15089-15094.

Yamazaki K, Oseto M, Seto Y, et al. Reverse transcription-polymerase chain reaction detection and sequence analysis of small round-structured viruses in Japan. Arch Virol Suppl 1996,12:271-276.

Yan B, Zemskova M, Holder S, et al. The PIM-2 kinase phosphorylates BAD on serine 112 and reverses BAD-induced cell death. J Biol Chem 2003,278:45358-45367.

Yang C, Li M, Newman RD, et al. Genetic diversity of HIV-1 in western Kenya: subtype-specific differences in mother-to-child transmission. Aids 2003,17:1667-1674.

Yang YH, Speed T. Design issues for cDNA microarray experiments. Nat Rev Genet 2002,3:579-588.

Yang Z. Estimating the pattern of nucleotide substitution. J Mol Evol 1994,39:105-111.
Yeung KY, Medvedovic M, Bumgarner RE. From co-expression to co-regulation: how many microarray experiments do we need? Genome Biol 2004,5:R48.

Youssef S, Stuve O, Patarroyo JC, et al. The HMG-CoA reductase inhibitor, atorvastatin, promotes a Th2 bias and reverses paralysis in central nervous system autoimmune disease. Nature 2002,420:78-84.

Yu X, Yu Y, Liu B, et al. Induction of APOBEC3G ubiquitination and degradation by an HIV-1 Vif-Cul5-SCF complex. Science 2003,302:1056-1060.

Zennou V, Petit C, Guetard D, Nerhbass U, Montagnier L, Charneau P. HIV-1 genome nuclear import is mediated by a central DNA flap. Cell 2000,101:173-185.

Zhang H, Dornadula G, Beumont M, et al. Human immunodeficiency virus type 1 in the semen of men receiving highly active antiretroviral therapy. N Engl J Med 1998,339:1803-1809.

Zhang H, Yang B, Pomerantz RJ, Zhang C, Arunachalam SC, Gao L. The cytidine deaminase CEM15 induces hypermutation in newly synthesized HIV-1 DNA. Nature 2003,424:94-98.

Zhang X, Zhao J, Li XD, et al. [Construction of adenovirus vector expressing TIP30 and its tumor suppressive effect in vitro and in vivo]. Zhonghua Zhong Liu Za Zhi 2004,26:85-88.

Zhang Y, Luxon BA, Casola A, Garofalo RP, Jamaluddin M, Brasier AR. Expression of respiratory syncytial virus-induced chemokine gene networks in lower airway epithelial cells revealed by cDNA microarrays. J Virol 2001,75:9044-9058.

Zhao H, Granberg F, Elfineh L, Pettersson U, Svensson C. Strategic attack on host cell gene expression during adenovirus infection. J Virol 2003,77:11006-11015.

Zheng YH, Plemenitas A, Fielding CJ, Peterlin BM. Nef increases the synthesis of and transports cholesterol to lipid rafts and HIV-1 progeny virions. Proc Natl Acad Sci U S A 2003,100:8460-8465.

Zhong P, Peeters M, Janssens W, et al. Correlation between genetic and biological properties of biologically cloned HIV type 1 viruses representing subtypes A, B, and D. AIDS Res Hum Retroviruses 1995,11:239-248.

Zhong XS, Zheng JZ, Reed E, Jiang BH. SU5416 inhibited VEGF and HIF-1alpha expression through the PI3K/AKT/p70S6K1 signaling pathway. Biochem Biophys Res Commun 2004,324:471-480.

Zhu H, Cong JP, Mamtora G, Gingeras T, Shenk T. Cellular gene expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. Proc Natl Acad Sci U S A 1998,95:14470-14475.

Zobiack N, Rescher U, Ludwig C, Zeuschner D, Gerke V. The annexin 2/S100A10 complex controls the distribution of transferrin receptor-containing recycling endosomes. Mol Biol Cell 2003,14:4896-4908.

Zou S. A practical approach to genetic screening for influenza virus variants. J Clin Microbiol 1997,35:2623-2627.

Chapter 7.0 Appendix

Table 1. HIV-1 subtype reference sequences

Accession	*New Key**	*Country of Origin*	*Published Subtype*
AF069670	25AAA1A1_36	Somalia	A
AF004885	26AAA1A1_25	Kenya	A
AJ251057	28AGAGAG_45	W/C.Africa	A/AG
M62320	27A1A1XX_42	Uganda	A1
AF286237	51AAA2A2_12	Cyprus	A2
AF286238	52AAA2A2_13	Congo	A2
AF193276	23ABABXX_30	Russia	AB
AF193277	24ABABAB_30	Russia	AB
U54771	33AEAEAE_41	Thailand	AE
U51189	34AEAEXX_41	Thailand	AE
U51188	35AEAEAE_9	Central African Republic	AE
AF197340	36AEAEAE_9	Central African Republic	AE/E?
L39106	29AGXXAG_29	Nigeria	AG
AF063223	30AGAGAG_14	Djibouti	AG
AF063224	31AGAGAG_14	Djibouti	AG
AJ251056	32AGAGAG_45	W/C.Africa	AG
AF049337	37Y1Y1XX_12	Cyprus	AGI
AF119819	38Y1Y1Y1_21	Greece/Cyprus	AGI
AF119820	39Y1Y1Y1_21	Greece/Cyprus	AGI
AJ245481	22Y1Y1Y1_27	Mali	AGIJ
AF064699	19Y1Y1Y1_8	Burkina Faso	AGJ
AJ288981	20Y1Y1Y1_35	Senegal	AGJ
AJ288982	21Y1Y1Y1_27	Mali	AGJ
NC_001802	12BBBBBB_0	None	B
U63632	13BBBBBB_44	USA	B
U21135	14BBBBBB_44	USA	B
M17451/ M12508	15BBBBBB_22	Haiti	B
AY008715	56BCBCBC_33	S.China	BC
AY008717	57BCBCBC_33	S.China	BC
AY008716	58BCBCBC_33	S.China	BC
AF385934	40BFBFBF_43	Uruguay	BF
AF385935	41BFBFXX_43	Uruguay	BF
AF385936	42BFBFXX_2	Argentina	BF
U52953	53CCXXCC_7	Brazil	C
AF286226	54BCBCBC_11	China	C
AF286230	55BCBCBC_11	China	C
AF110967	59CCCCCC_6	Botswana	C
AF067155	60CCCCCC_23	India	C
U46016	61CCCCCC_15	Ethiopia	C
AF289548	8CDCDCD_40	Tanzania	CD
AF289550	9CDCDCD_40	Tanzania	CD

AF289549	11CDCDCD_40	Tanzania	CD
AF179368	18Y1XXXX_20	Greece	cpx
AJ291719	17Y1Y1Y1_16	France	CRF
M22639 (U16633, Z2)	3DDDDDD_0	none	D
KO3454/X04414	4DDXXDD_13	Congo	D
M27323	5DDDDDD_13	Congo	D
U88824	10DDXXDD_42	Uganda	D
AF005494	43FFF1F1_7	Brazil	F
AF077336	44FFF1F1_13	Congo	F
AF076998	6DFDFDF_13	Congo	F/FD
AF193253	7DFDFDF_13	Congo	F/FD
AF005495	16BBXXXX_7	Brazil	F1
AF075703	45FFF2F1_25	Kenya	F1
AJ249238	46FFF1F1_1	Africa	F1
AJ249237	47F2F2F2_10	Cameroon	F2
AJ249239	49KKKKKK_10	Cameroon	F3
AJ249235	50KKXXKK_13	Congo	F3
AF061640	62GGGGGG_25	Kenya	G
AF061642	63GGGGGG_13	Congo	G
AF084936	64GGGGGG_13	Congo	G
AF005496	65HHHHHH_9	Central African Republic	H
AF190128	66HHXXHH_4	Belgium	H
L11793	67HHXXXX_13	D.R. of Congo	H
AF190127	68HHHHHH_4	Belgium	H
AF082394	69JJJJJJ_13	Congo	J
AF082395	70JJJJJJ_13	Congo	J
AJ249236	48F2XXF2_10	Cameroon	K
AJ006022	1NNNNNN_10	Cameroon	N
AJ271370	2NNNNNN_10	Cameroon	N
L20587	71OOXXXX_10	Cameroon	O
L20571	72OOXXXX_10	Cameroon	O

* Refer to results section 1.3.2 for full explanation of new key

Table 2. Table showing manual subtype classification of full length Los Alamos reference sequences and NCBI HIV genomes.

Classification was performed by three separate methods: maximum parsimony (Phylip, protpars), neighbour joining (ClustalW) and agglomerative hierachical clustering using complete linkage based on sequence identity matrices. For each genome region, classification is given as both group and subgroup (within each tree). Sequence accessions, STAR Identifier, whether the sequence is a reference strain, geographical origin and published subtype, are all noted.

SEE EXCEL FILE Table2.xls ON ENCLOSED CD

Table 3. Gene lists.

Gene lists for: **(Worksheet A)** Significantly up and down regulated genes in response to HIV-1 subtype B at 24 HPI; **(Worksheet B)** in response to HIV-1 group O at 24 HPI; **(Worksheet C)** in response to HIV-2 ETP at 24 HPI. Both gene ID (accession number) and name are given.

SEE EXCEL FILE Table3.xls ON ENCLOSED CD

Figure 1. FACS analysis of forty CEM-G clones, infected with CCR5-lentivirus and grown out under selection, after limiting dilution cloning.

Each line is stained for CCR5 (FL2-red) and background level of green fluorescence detected (FL1-green).

SEE SEPARATE EXCEL FILE (Figure 1.xls)